

# **Theory of Optimal Experiments**

# **Probability and Mathematical Statistics**

*A Series of Monographs and Textbooks*

## *Editors*

**Z. W. Birnbaum**

University of Washington  
Seattle, Washington

**E. Lukacs**

Catholic University  
Washington, D C

- 
- 1 Thomas Ferguson Mathematical Statistics A Decision Theoretic Approach 1967
  - 2 Howard Tucker A Graduate Course in Probability 1967
  - 3 K. R Parthasarathy Probability Measures on Metric Spaces 1967
  - 4 P Revesz The Laws of Large Numbers 1968
  - 5 H P McKean, Jr Stochastic Integrals 1969
  - 6 B V Gnedenko Yu K Belyayev, and A D Solov'yev Mathematical Methods of Reliability Theory 1969
  - 7 Demetrios A Kappos Probability Algebras and Stochastic Spaces 1969, 1970
  - 8 Ivan N Pesin Classical and Modern Integration Theories 1970
  - 9 S Vajda Probabilistic Programming 1972
  - 10 Sheldon Ross Introduction to Probability Models 1972
  - 11 Robert B Ash Real Analysis and Probability 1972
  - 12 V V Fedorov Theory of Optimal Experiments 1972
  - 13 K V Mardia Statistics and Directional Data 1972

## *In Preparation*

Tatsuo Kawata Fourier Analysis in Probability Theory

Fritz Oberhettinger Tables of Fourier Transforms and Their Inverses

H Dym and H P McKean Fourier Series and Integrals

FACULTY OF ENGINEERING LIBRARY  
THE UNIVERSITY OF JODHPUR

Acc. No.....

Call No.....

# THEORY OF OPTIMAL EXPERIMENTS

V. V. Fedorov

*Moscow State University  
Interfaculty Laboratory of Statistical Methods  
Lenin Hills, Moscow*

*Translated and Edited by*

*W. J. Studden and E. M. Klimko*

DEPARTMENT OF STATISTICS  
PURDUE UNIVERSITY  
LAFAYETTE, INDIANA

COPYRIGHT © 1972 BY ACADEMIC PRESS INC

ALL RIGHTS RESERVED

NO PART OF THIS BOOK MAY BE REPRODUCED IN ANY FORM  
BY PHOTOSTAT MICROFILM RETRIEVAL SYSTEM OR ANY  
OTHER MEANS WITHOUT WRITTEN PERMISSION FROM  
THE PUBLISHERS

ACADEMIC PRESS, INC.  
111 Fifth Avenue New York, New York 10003

*United Kingdom Edition published by*  
ACADEMIC PRESS INC (LONDON) LTD  
24/28 Oval Road London NW1 7DD

LIBRARY OF CONGRESS CATALOG CARD NUMBER 76 182610

AMS (MOS) 1970 Subject Classification 62 02

PRINTED IN THE UNITED STATES OF AMERICA

Originally published in Russian under the title  
TEORIYA OPTIMAL NOGO EKSPERIMENTA  
by  
Izdatelstvo Moskovskogo Universiteta 1969

# Contents

<i>Foreword</i>	ix
<i>Preface</i>	xi

<b>Introduction</b>	1
---------------------	---

## Chapter 1—Regression Analysis and Optimality Criteria for Regression Experiments

1.1 Basic Elements of Matrix Algebra	12
1.2 General Requirements Satisfied by Estimates	21
1.3 Best Linear Estimates	23
1.4 The Search for Estimates for Nonlinear Parametrization. Best Quasi- Linear Estimates	33
1.5 Estimation of the Dispersion of the Resulting Observations. The Efficiency of an Experiment	36
1.6 Regression Analysis in the Presence of Errors in the Determination of the Control Variables	40
1.7 Analysis of Experimental Data in the Case of Simultaneous Observations of Several Variables	49
1.8 Methods of Comparing Results of Experiments	51
1.9 The Loss Function for Regression Experiments	56
1.10 The Concept of Experimental Design. Continuous Normalized Designs	58

## Chapter 2—Continuous Optimal Designs (Statistical Methods)

2.1 Basic Properties of the Information Matrix	64
2.2 Equivalence of $D$ -Optimal and Minimax Designs. Basic Properties of These Designs	68
2.3 One-Dimensional Polynomial Regression	83
2.4 Trigonometric Regression on an Interval	94

2 5	Computational Methods for Construction of <i>D</i> Optimal Designs	97
2 6	Some Particular Iterative Procedures for Constructing <i>D</i> Optimal Designs	104
2 7	Truncated <i>D</i> Optimal Designs	114
2 8	Nonlinear Parametrization of a Response Surface Local <i>D</i> Optimal Designs	120
2 9	Linear Criteria of Optimality	122
2 10	Iterative Methods for Constructing Linear Optimal Designs	132
2 11	Designs Minimizing $\text{Tr } D(\epsilon)$	137
2 12	Designs Minimizing the Mean Dispersion of the Estimate of the Response Surface over the Domain of Values	142
2 13	Extrapolation at a Point	145
2 14	Quadratic Loss	153

### **Chapter 3—Properties and Methods of Construction for Optimal Discrete Designs**

3 1	Discrete Designs	155
3 2	Properties and Methods of Constructing <i>D</i> -Optimal Designs	160
3 3	Construction of Discrete Linear Optimal Designs	167

### **Chapter 4—Sequential Methods of Designing Experiments for Refining and Determining Estimates of the Parameters**

4 1	Some Generalities of Contemporary Experimental Investigations	171
4 2	Sequential <i>D</i> Optimal Design (Linear Parametrization and Time Constant Efficiency of the Experiment)	173
4 3	Sequential Linear Optimal Designs (Linear Parametrization and Time Constant Efficiency of the Experiment)	183
4 4	Sequential Designs for Nonlinear Parametrization	186
4 5	Designs When the Efficiency Function of the Experiment Is Unknown	199
4 6	Design in the Presence of Errors in the Determination of the Control Variables	202
4 7	Construction of Optimal Designs When the Experimental Conditions Vary in Time	203

### **Chapter 5—Design of Experiments in the Case of Simultaneous Observation of Several Random Quantities**

5 1	Basic Properties of the Information Matrix	209
5 2	<i>D</i> Optimal Designs	211
5 3	Linear Optimal Designs	221
5 4	Sequential Design	224

**Chapter 6—Discriminating Experiments**

6.1 Statement of the Problem	226
6.2 Criteria Depending on the Difference of Sums of Weighted Squares of Residuals	231
6.3 The Method of Likelihood Ratio	249
6.4 Discriminating Hypotheses Based on the Entropy Measure of Information	257

**Chapter 7—Generalized Criteria of Optimality**

7.1 Experiments Minimizing Generalized Loss	264
7.2 The Information Approach to the General Problem of Seeking the True Mathematical Model	267

<b>References</b>	285
-------------------	-----

<i>Subject Index</i>	289
----------------------	-----

## **Foreword**

This book is a translation of the Russian monograph "Teoriya Optimal'nogo Eksperimenta" by V.V. Fedorov. The text is concerned with the problem of optimal allocation of resources in conducting experiments (the optimal design of experiments). It incorporates a considerable amount of new material, which, up to now, was scattered throughout the literature. Much of this research has been done by the author. For this reason the translators feel that this work should be made available in the English literature on the subject. The book should be of interest not only to statisticians, but also to anyone using regression analysis in his work.

We would like to thank Professor V. V. Fedorov for supplying us with an updated copy of his text. In addition various minor errors have been corrected.

The translators would like to thank the Department of Statistics of Purdue University for generous encouragement, support, and secretarial help in the preparation and typing of the manuscript. The work was also partially supported by the National Science Foundation under grant number GP 20306.

W. J. STUDDEN  
E. M. KLIMKO

## Preface

Experimental design (to be more accurate, regression experimental design) is a comparatively new and rapidly developing branch of mathematical statistics. In this book, an attempt was made to present the mathematical apparatus of regression experimental design so that it would be accessible to comparatively broad circles of researchers and technologists.

I would like to express my sincere gratitude to Professor W. J. Studden and Professor E. M. Klimko who have undertaken a labor-consuming task of translating the preprint of the book, which contained a number of misprints and inaccuracies. I am flattered by the fact that one of the translators of the book is Professor W. J. Studden who has contributed so much to the mathematical theory of regression experimental design.

V. V. FEDOROV

# Introduction

I. At the present level of development of science and technology, many investigations in physics, biology, chemistry, metallurgy, etc., require setting up complicated and expensive experiments. The measurement of any experimental quantity always takes place under the influence of some obstacles which can never be completely eliminated despite the efforts of the researcher to keep them to a minimum. Because of this, the investigator deals not with deterministic, but with random quantities. In some cases the measured quantities are random by their very nature. It is necessary to deal with the measurement of such quantities in quantum mechanics, in biological investigations, in certain problems of chemical kinetics, and in other branches of science.

The necessity of applying the apparatus of mathematical statistics to the reduction of the results of measurements is evident when the random components are commensurate with the results themselves. The corresponding methods of reduction have long been used in experimental practice.

For a long time, the attention of mathematical statistics was focused on the perfection of methods of reduction when the method of conducting the experiment was preestablished. The choice of the

experiment itself, that is, when and where to take measurements, was determined mainly by the intuition of the experimenter. During this time it was necessary to deal with problems which were comparatively simple from the theoretical and experimental viewpoints, and which did not require significant expenditures (financial means, time, limited material resources). Losses related to errors of the intuitive solution for the method of conducting the experiment were not met very often and were not essential.

The development of science and technology led to natural complications in the theoretical interpretation of the results obtained, and in the methods of carrying out necessary experimental investigations. More complicated experimental situations led to sharply increased cost of experimental investigations. As an example, one may cite investigations in the realm of the physics of elementary particles where the necessity of building powerful accelerators makes measurements very expensive. Therefore, the problem of extracting an increased quantity of data from processes under study with finite resources is currently very real. Relying on the intuition of the experimenter for the solution of a given problem becomes less and less hopeful. In connection with this, it is absolutely necessary to give a broad class of methods which would give not only the means of reduction of experimental data, but also would permit the organization of the experiment in an optimal manner.

The mathematical apparatus used in the optimal organization of experiments is based on a composition of mathematical statistics methods and methods of solving extremal problems. Increasingly often, mathematical statistics is necessary for wise construction and elucidation of the basic properties of the criteria of optimality of an experiment. Afterward the problem of optimal organization of an experiment (or more briefly, a design of an experiment) leads to the solution of some extremal problem.

We note that a design is suitable only in those cases when the experimenter clearly sets forth the end-purpose of the investigation being conducted. It may be further added that statistical methods of design are instruments which make attainment of an established goal easier. Moreover, the effectiveness of utilizing any instrument (or apparatus) essentially depends on how well it is used and on the qualifications of the hands using it. In the same way, the effectiveness of applying experimental design methods depends, in the large, on how well they are mastered, and on their appropriate utilization. For example, in

conducting elementary experiments requiring very few resources, it is hardly necessary to apply methods requiring computational resources which could significantly exceed the cost of the experiment itself.

II. Currently, it is possible to divide the mathematical theory of experimental design into two basic areas: design of extremal experiments and design of experiments for an elucidation of the mechanism of a phenomenon. A design of the first type is used in those cases when the experimenter is interested in conditions under which the process being investigated satisfies some criteria of optimality. For example, in the development of a new chemical-technological process, the criteria of optimality consists of requiring maximal output of the products of the reaction. In this case, design consists of finding those values of temperature, pressure of reagents, their percentage of concentration, and so forth, for which the established requirements are satisfied.

Frequently, the experimenter finds it necessary to elucidate the global behavior of an investigated object, or as we shall say in the future, to elucidate the mechanism of a phenomenon. For example in the study of a chemical-technological process it may be necessary to elucidate the dependence of the final products of the reaction on temperature, pressure, reagents, and so forth. In the language of mathematics a similar type of problem is formulated in the following way: It is necessary to find a function which defines the relationship between the end product of the reaction and the quantities introduced at the beginning of the reaction (temperature, percentage of reagent concentration, and so forth). Or in short: Find a mathematical model of the given process. In this way, by an elucidation of the mechanism of a phenomenon, here, in distinction from the usual use of this term, it will be understood to be not a direct investigation of a reaction on the level of elementary particles, molecules and so on, but an investigation of the phenomenological side of an event. In other words, we are indifferent, for example, to the manner in which two molecules react. The important point to us is the dependence of the final product on the percentage of reagent concentration which enters into this reaction and can be directly measured in a given experiment. Having set up an investigation of the mathematical form of the dependence of some quantity on corresponding factors, we will thereby give enough information to the specialist of a branch of science on the basis of which he may call on the necessary theoretical apparatus and be able

to deduce the correct form of the elementary chemical reactions

The design of an extremal experiment has been adequately investigated. A significant number of articles and surveys have been devoted to an exposition of the corresponding mathematical apparatus. The results of a majority of these are discussed in the book by Nalimov and Chernova [1]. The book of Hicks, "Fundamental Concepts in the Design of Experiments" [2], should also be pointed out. This book is usually used in a first introduction to the problems under consideration. Along with the broad literature expounding the mathematical aspects of design of extremal experiments there are many works related to the practical application of methods of designing such experiments (cf., for example, [3, 4]).

Essentially little attention in contemporary literature is given to the design of experiments for seeking mathematical models to describe an investigated object. From the Russian literature here we can cite only the monograph of Klepikov and Sokolov [5], in which there is a brief section related to the planning of such experiments, and a chapter of Nalimov [4]. By comparison with the number of works on extremal design, only a very small collection of works deal with the practical application of methods of experimental design for seeking mathematical models.

This book sets as its aim the presentation, from the viewpoint of practical applications, of the more important and accessible mathematical methods of experimental design, along the line of defining mathematical models which describe an investigated object. For a clear explanation of the possibilities and the region of applicability of each of the methods considered in this volume, theoretical material is accompanied by a significant number of examples.

III. We now consider a more detailed mathematical setup of the problem of experimental design along the line of elucidating the mechanism of a phenomenon. Usually the measured quantity depends on one or several factors, which sometimes we shall call "control variables," this term attempts to emphasize those values of each of the factors that can be chosen arbitrarily from some given domain. Various quantities can be considered as control variables, depending on the type of experiment. For example time, scattering angle of particles falling on a target, temperature, tension, output of an experimental apparatus, percentage of reagent concentration in a chemical or biological experiment, and so on.

The column vector

$$\mathbf{x} = \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{vmatrix}$$

is set up for each level of these quantities; the coordinates  $x_1, x_2, \dots, x_k$  are equal to the values of the control variables and are enumerated in an order suitable to the experimenter.

The space of dimension  $k$ , on which the vector  $\mathbf{x}$  is defined, is called the factor space or the space of control variables. The collection of points of this space, where measurements are possible (that is, the corresponding values of the control variables  $x_1, x_2, \dots, x_k$  which can be realized by the experimenter) is called the region of possible measurements or the domain of actions, and in this book is indicated by  $X$ . Determination of the boundary of the region  $X$  plays an important role in the design of optimal experiments. In some cases, these boundaries are defined by the very nature of the control variables. For example, the percentage of reagent concentration cannot be less than 0 or greater than 100 %, the dimensions of details under investigation are negative, and so on. In other cases, met significantly more often, the boundary of the domain of action is determined by the characteristics of the apparatus used by the investigator or by the form of the process under study. Indeed, the upper boundary of temperature will be determined either by the power of the heat source or by the thermal insulation properties of the materials. The upper limits for the speed of accelerated elementary particles is determined by the parameters of the accelerator. The reader can extend without difficulty the list of analogous examples, related to real experiments encountered in the branch of science familiar to him.

The problem of setting up an experiment for finding a mathematical model, as already noted, is the search for relations among the measured quantities (sometimes there can be several) and the control variables. Since the results of an observation are random quantities, it makes sense in the majority of cases to talk about the relation of the mean value of the quantity under study to the control variables. In the future we shall assume that this relation can be written by means of some function

$$E(y | x) = \eta(x),$$

where  $E(y | x)$  is the mean value of the quantity  $y$  under study for

values of the control variables defined by means of the coordinates of the vector  $x$ . The function  $\eta(x)$  depends on unknown parameters  $\theta_1, \theta_2, \dots, \theta_m$ , and in the general case, its analytic form can also be unknown.

Beginning the search for a mathematical model [the function  $\eta(x)$ ], the experimenter possesses some prior information. The degree of this information can be characterized at three basic levels:

1. The analytic form of the function  $\eta(x) = \eta(x, \theta)$  is known. It is required to determine or estimate the unknown parameters.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$$

2. It is known that the analytic form of the function is defined by one of the functions

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

The dimension of the vectors  $\theta_1, \theta_2, \dots, \theta_r$  can be different. It is required to determine which of the functions

$$\eta_1(x, \theta_1), \eta_2(x, \theta_2), \dots, \eta_r(x, \theta_r)$$

is correct and to find the unknown parameters.

3. The analytic form of the function  $\eta(x, \theta)$  is not known. It is known only that the function  $\eta(x)$  can be approximated sufficiently well, in the region of interest to the experimenter, by means of a finite series in some system (or systems) of given functions. It is required to find the best description of the function  $\eta(x)$ .

Although the decomposition 1-3 is sufficiently coarse, it is possible to find examples where the real experimental situation occupies a position between any two of these levels. The levels 1-3 are convenient from the viewpoint of existing methods of designing an experiment, and, at the same time, they describe fairly well the majority of real cases.

The mathematical apparatus of experimental design with prior

knowledge corresponding to level 1 began to develop about in the mid 1950s, and its development has practically been completed. For this case, the development consists of effective methods of statistical and sequential experimental design. By a statistical design of an experiment, here we understand a prior design of an experiment in its entirety. By a sequential design we mean a design of an experiment in stages. That is, one or several measurements is taken; after these measurements are realized a reduction of the obtained data is carried out and planning of a further stage is begun, and so on. For a broad class of functions  $\eta(x, \theta)$ , a statistically designed experiment consists essentially in making use of prepared tables describing the characteristics of optimal designs.

Methods of designing experiments for seeking the correct model from some given collection of models (level 2) appeared only recently and undoubtedly will still improve both from the conceptual as well as from the computational point of view. In connection with this fact, basic attention is focused here on the more simple and complete methods. For more complicated or less complete methods, from the theoretical point of view, the presentation of the material takes on a more descriptive character. The majority of the methods considered for designing experiments, with prior knowledge corresponding to level 2, are by their very nature sequential. It is to be pointed out that the given methods of planning experiments are more effective when the number of concurring models  $\eta_1(x, \theta_1), (x, \theta_2), \dots, \eta_r(x, \theta_r)$  is small. This is a statement of the fundamental regularity of design in general: The more we know, the more effectively we can plan. Therefore the problem of the experimenter as a specialist in his branch of science, which gives rise to the necessity of conducting a given investigation, is to seek the smallest possible collection of models based on a careful analysis of available theoretical and experimental data.

A more difficult, less worked out, and very often met problem is the design of experiments when the analytical function  $\eta(x, \theta)$  is completely unknown (level 3). Generally, it is hardly possible to design an experiment which would permit the formulation of the problem in its entirety. Nevertheless its solution can be reduced to some sequential procedure, which contains alternating experiments (design and practical realization) of the following form:

1. The functional form  $\eta = \eta(x, \theta)$  is known. It is required to determine or estimate the parameter  $\theta$ .

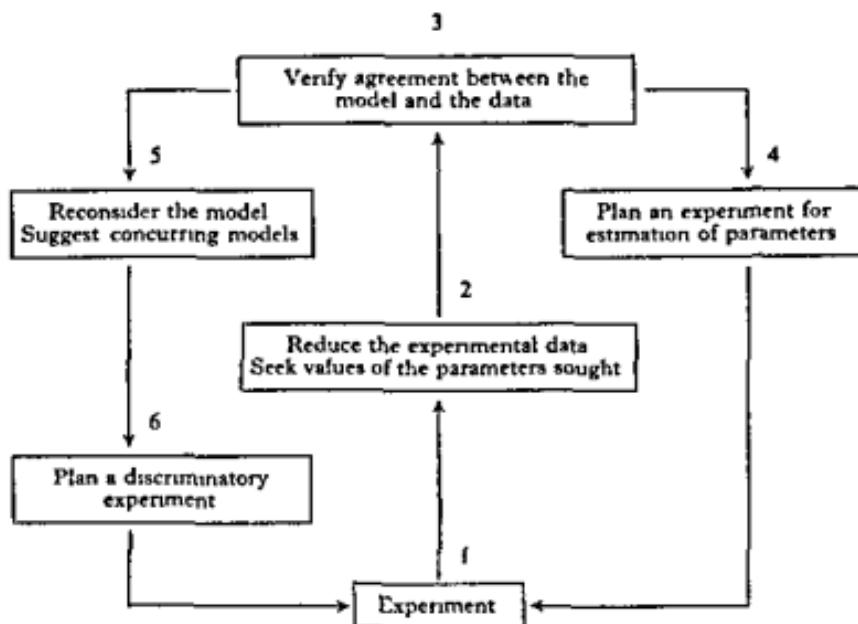
2 On the basis of a theoretical analysis of the occurring hypotheses, or from the results of previous experiments, two or several hypotheses are suggested about the form  $\eta(x)$

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

It is required to find the dependence  $\eta_j(x, \theta_j)$ , this means the best form describing the object under study

A more detailed sequential process for seeking a mathematical model is presented in Diagram 1

Diagram 1



Block 1 corresponds to the experimental stage of the work, i.e., the technical carrying out of previously designed experiments. Usually the carrying out of designed experiments is preceded by conducting some "preliminary" experiment which offers the experimenter rough information about the process under investigation, since in the complete absence of prior information, design is impossible.

The next stage of the work (Block 2) is the computation of estimates

of the parameters  $\theta$  under the assumption that the functional form of  $\eta(x, \theta)$  is known. Sometimes the computation of the estimates of the parameters is preceded by an analysis of experimental data from the point of view of discrimination of concurring models.

After estimates of the parameters are found, it is necessary to verify whether or not the behavior of the function  $\tilde{\eta}(x) = \eta(x, \tilde{\theta})$ , where  $\tilde{\theta}$  is the value of the estimate, agrees with the experimental data. (Block 3).

If  $\tilde{\eta}(x)$  satisfies the experimental data sufficiently well, then, depending on the circumstances of the experiment, the experimenter must either stop or design a supplementary experiment for estimating the entire collection of parameters or some group of them which are of more interest to the experimenter (Block 4).

If  $\tilde{\eta}(x)$  does not satisfy the experimental data, then the necessity of a more careful analysis of the occurring phenomena arises. In the absence of any reduction in the number of suggested models we must turn to the design of a more precise experiment. If we have facts which point out the possibility of describing the phenomena under study by some other model in comparison with the original model, then it is necessary to begin to design an experiment which would permit us to clarify which of these models best describes the objects under study (Block 6). In this way, the strategy for conducting an experiment to elucidate the mathematical model, with prior knowledge corresponding to level 3, can be represented in the form of a sequence of cycles 4-1-2-3- and 5-6-1-2-3 (cf. Diagram 1). The order of alteration of these cycles is determined by the results of verifying agreement between the model and the data (Block 3). In many cases, an analogous strategy can be applied to experimental design corresponding to the second level of prior knowledge.

IV. We now go to a brief description of the contents of the following chapters. In Chapter 1, a survey of regression analysis is presented at a level necessary for the construction of the mathematical apparatus of experimental design. The concept of best linear estimator (or best quasi-linear estimator, for the case of functions  $\eta(x, \theta)$  depending nonlinearly on the parameters  $\theta$ ) is presented as fundamental in constructing the scheme of regression analysis.

Such an approach permits us to develop the theory of optimal experiments for determining the estimator of an unknown parameter not depending on a concrete form of the distribution function of

the observations. The first section of this chapter has the character of a review, and contains a survey of matrix algebra. Its aim is to gather in one place the basic formulas of matrix algebra, necessary for regression analysis and design of experiments, which are scattered throughout various sources, and at the same time, make it easy for the reader to become acquainted with the basic material. The concluding section of the first chapter is devoted to the formulation of basic optimality criteria of experiments for determining and estimating unknown parameters. The formal definition of an experimental design is introduced there.

Chapter 2 is devoted to an exposition of the basic properties of continuous statistically optimal designs for the various criteria of optimal design and for various criteria of optimality. The analytical and computational methods of constructing such designs is investigated here. A table of characteristics of optimal designs is also presented.

The properties and methods of construction of optimal designs, taking into account the discrete character of the expenditures for carrying out real experiments, is studied in Chapter 3.

In Chapter 4, results are presented for sequential methods of designing experiments in the determination and estimation of unknown parameters. These methods, above all, give perspective to the non-linear parametrization functions  $\eta(x, \theta)$  and in the highest degree satisfy the spirit of Diagram 1.

Chapter 5 generalizes the results obtained in the previous chapters to the case where it is possible to have simultaneous measurements, some of which generally speaking are correlated.

Chapter 6 contains information on the discrimination of statistical hypotheses, as they apply to problems of seeking the correct mathematical model and on various methods of designing discriminating experiments. In contrast to the results of the preceding five chapters, Chapter 6 depends essentially on the analytic form of the distribution function of the observations.

In the concluding chapter, methods of experimental design depending on generalized criteria of optimality are considered which succeed in simultaneously solving the problems of deciding on the correct mathematical model and estimating unknown parameters. Basic attention is focused on criteria using the entropy measure of information.

V. The material is presented in such a way that the basic theoretical results are formulated in the form of theorems. Besides making it possible to concentrate on the most important results, theorems permit the reader to become acquainted with the mathematical methods of designing optimal experiments. He may confine himself, at a first reading, only to the analysis contained in the theorems and explanations, skipping the more complicated and cumbersome proofs. Theorems and lemmas are enumerated in the following form: The first digit is the chapter number, the second digit is the section number, and the last digit is the ordering number in the given section; formulas are handled by means of analogous numbering.

# 1

## Regression Analysis and Optimality Criteria for Regression Experiments

### 1.1. Basic Elements of Matrix Algebra

This section is of a review character. Detailed illustrations, definitions, and proofs of the basic results presented in what follows can be found in [6-8].

#### I. Basic properties of matrices

**DEFINITION 1** The rectangular array of numbers

$$\begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{vmatrix} \quad (1.1)$$

is called a matrix.

If  $m = n$ , then the matrix is called square, and the number  $m$ , equal to  $n$ , is called its order. If  $m = 1$  or  $n = 1$ , then the matrix is respectively a row vector or a column vector. In general, matrices are called rectangular (with dimension  $m \times n$ ) or  $m \times n$  matrices. The numbers comprising the matrix are called its elements.

In the double index notation on the elements, the first index is called the row number, and the second index is called the column number. When  $m = 1$  or  $n = 1$  one of the indices is omitted. We will indicate the  $m \times n$  matrix by one symbol  $A$  or

$$\| A_{ik} \| \quad (i = 1, 2, \dots, m; k = 1, 2, \dots, n). \quad (1.1.2)$$

If  $A$  is a square matrix, then instead of (1.1.2) the notation  $\| A_{ik} \|_1^n$  is used.

**DEFINITION 2.** The matrix composed of part of the elements of  $A$  is called a submatrix of the matrix  $A$ .

**DEFINITION 3.** The rectangular matrix  $A'$  of dimension  $n \times m$  is called the transpose of the matrix  $A$  of dimension  $m \times n$ , if  $A'_{ik} = A_{ki}$  ( $i = 1, 2, \dots, n; k = 1, 2, \dots, m$ ).

It is easy to verify the following properties:

- (1)  $(A + B)' = A' + B'$ ;
- (2)  $(\alpha A)' = \alpha A'$ ;
- (3)  $(AB)' = B'A'$ ;
- (4)  $(A')' = A$ .

**DEFINITION 4.** The sum of two rectangular matrices  $A$  and  $B$  both of dimension  $m \times n$  is the matrix  $C$  of the same dimensions, the elements of which are equal to the sums of the corresponding elements of the matrices  $A$  and  $B$ :  $C = A + B$ , if

$$C_{ik} = A_{ik} + B_{ik} \quad (i = 1, 2, \dots, m; k = 1, 2, \dots, n). \quad (1.1.3)$$

From Definition 4 it follows immediately that

- (1)  $A + B = B + A$ ;
- (2)  $(A + B) + C = A + (B + C)$ .

Here  $A, B, C$  are arbitrary rectangular matrices of the same dimensions.

**DEFINITION 5.** The product of a matrix  $A$  by a number  $\alpha$  is the matrix  $C$ , the elements of which are obtained from the corresponding elements of the matrix  $A$  multiplied by  $\alpha$ :  $C = \alpha A$ , if  $C_{ik} = \alpha A_{ik}$  ( $i = 1, 2, \dots, m; k = 1, 2, \dots, n$ ).

It is easy to see that

- (1)  $\alpha(A + B) = \alpha A + \alpha B$ ;
- (2)  $(\alpha + \beta)A = \alpha A + \beta A$ ;
- (3)  $(\alpha\beta)A = \alpha(\beta A)$ .

The difference of two matrices is defined by the equality

$$A - B = A + (-1)B \quad (1.14)$$

**DEFINITION 6** The product of two rectangular matrices

$$A = \begin{vmatrix} A_{11} & A_{12} & A_{1n} \\ A_{21} & A_{22} & A_{2n} \\ A_{m1} & A_{m2} & A_{mn} \end{vmatrix} \quad \text{and} \quad B = \begin{vmatrix} B_{11} & B_{12} & B_{1q} \\ B_{21} & B_{22} & B_{2q} \\ B_{n1} & B_{n2} & B_{nq} \end{vmatrix}$$

is the matrix

$$C = \begin{vmatrix} C_{11} & C_{12} & C_{1q} \\ C_{21} & C_{22} & C_{2q} \\ C_{m1} & C_{m2} & C_{mq} \end{vmatrix},$$

the elements of which are equal to

$$C_{ij} = \sum_{k=1}^n A_{ik}B_{kj} \quad (i = 1, 2, \dots, m \quad j = 1, 2, \dots, q) \quad (1.15)$$

From Definition 6 one can immediately show that

- (1)  $(AB)C = A(BC)$ ,
- (2)  $(A + B)C = AC + BC$ ,
- (3)  $A(B + C) = AB + AC$

We note that, in general,  $AB \neq BA$

**DEFINITION 7** The square matrix  $A^{-1}$  of order  $m$  is called the inverse of the matrix  $A$ , if

$$AA^{-1} = A^{-1}A = I_m, \quad (1.16)$$

where  $I_m = \{ \delta_{ik} \}_{1^m}$  is a square matrix whose diagonal elements are equal to one and the remainder are equal to zero

From Definition 7 it follows that

- (1)  $(AB)^{-1} = B^{-1}A^{-1}$ , if  $A$  and  $B$  are square matrices of the same dimension,
- (2)  $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$ ,
- (3)  $(A')^{-1} = (A^{-1})'$ ,
- (4)  $(A^{-1})^{-1} = A$

## II. The determinant of a square matrix.

**DEFINITION 8.** The determinant of the  $m$ th order or the determinant of the square matrix  $A$  of order  $m$  is the number obtained from the elements of the matrix according to the formula

$$|A| = |A_{m \times m}| = \begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{vmatrix} = \sum (-1)^k A_{1k_1} A_{2k_2} \dots A_{mk_m}, \quad (1.1.7)$$

where  $k_1, k_2, \dots, k_m$  run through all possible  $m!$  permutations of the numbers  $1, 2, \dots, m$ ;  $k$  is equal to the number of inversions in each permutation.

The determinant composed of part of the elements of the matrix  $A$  is indicated in the following form:

$$A \left( \begin{matrix} i_1, i_2, \dots, i_p \\ k_1, k_2, \dots, k_p \end{matrix} \right) = \begin{vmatrix} A_{i_1 k_1} & A_{i_1 k_2} & \cdots & A_{i_1 k_p} \\ A_{i_2 k_1} & A_{i_2 k_2} & \cdots & A_{i_2 k_p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i_p k_1} & A_{i_p k_2} & \cdots & A_{i_p k_p} \end{vmatrix}. \quad (1.1.8)$$

From the definition of the determinant:

- (1)  $|A'| = |A'|$ ;
- (2)  $|A| = |A^{-1}|^{-1}$ ;
- (3)  $|\alpha A| = \alpha^m |A|$ ,

where  $m$  is the order of the matrix  $A$ .

**DEFINITION 9.** If the determinant of the matrix is equal to zero, then the matrix is called singular.

**DEFINITION 10.** The rank of an arbitrary matrix  $A$  is the dimension of the largest square submatrix whose determinant is distinct from zero.

**Theorem 1.1.1 (Binet-Cauchy Formula).** If  $C = AB$ , where  $A$  is a matrix of dimension  $m \times n$  and  $B$  is of dimension  $n \times m$ , then if  $n \leq m$ ,

$$|C| = \sum_{1 \leq k_1 < k_2 < \dots < k_m \leq n} A \left( \begin{matrix} 1, 2, \dots, m \\ k_1, k_2, \dots, k_m \end{matrix} \right) B \left( \begin{matrix} k_1, k_2, \dots, k_m \\ 1, 2, \dots, m \end{matrix} \right). \quad (1.1.9)$$

From (1.1.9), it follows, for example, that

$$|AB| = |A| \cdot |B| \quad (1.1.10)$$

If  $C = AB$  and  $r_A, r_B, r_C$  are the ranks of the matrices  $A, B, C$ , then from (1.1.9)

$$r_C \leq \min(r_A, r_B) \quad (1.1.11)$$

**Theorem 1.1.2.** Let the matrix  $A$  have dimension  $m \times n$  and the matrix  $B$  be diagonal ( $B_{ik} = 0$ , if  $i \neq k$ ,  $1 \leq i, k \leq m$ ), then

$$|ABA| = \sum \left[ A^2 \begin{pmatrix} 1 & 2 & \dots & m \\ i_1 & i_2 & \dots & i_m \end{pmatrix} \prod_{a=1}^m B_{i_a i_a} \right] \quad (1.1.12)$$

where the sum extends over all possible minors composed of  $m$  rows and  $m$  columns numbered  $i_1, i_2, \dots, i_m$

### III Operations with partitioned matrices

We decompose the matrix  $A$  [cf (1.1.1)] into rectangular blocks

$$A = \begin{array}{ccc} \overbrace{A_{11}}^{n_1} & \overbrace{A_{12}}^{n_2} & \overbrace{A_{1t}}^{n_t} \\ \overbrace{A_{21}}^{n_1} & \overbrace{A_{22}}^{n_2} & \overbrace{A_{2t}}^{n_t} \\ \overbrace{A_{t1}}^{n_1} & \overbrace{A_{t2}}^{n_2} & \overbrace{A_{tt}}^{n_t} \end{array} \begin{array}{c} } \\ m_1 \\ } \\ m_2 \\ } \\ m_s \end{array}$$

Operations on partitioned matrices are carried out according to the same formal rules as in the case when there are numerical elements instead of blocks (cf Part I). Because of this, the decomposition into blocks must be such that the operations  $A_{ij} + B_{ij}$  and  $A_{ik} - B_{ik}$  make sense

**Theorem 1.1.3** Let

$$A = \begin{vmatrix} A & B \\ C & D \end{vmatrix}$$

then

$$A = |A| |D - CA^{-1}B| = |A - BD^{-1}C| |D| \quad (1.1.13)$$

**Theorem 1.1.4 (Frobenius Formula)** Let

$$M = \begin{vmatrix} A & B \\ C & D \end{vmatrix}$$

then

$$M^{-1} = \begin{vmatrix} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ -H^{-1}CA^{-1} & H^{-1} \end{vmatrix} \quad (1.1.14)$$

if  $|A| \neq 0$ , and

$$M^{-1} = \begin{vmatrix} K^{-1} & -K^{-1}BD^{-1} \\ -D^{-1}CK^{-1} & D^{-1} + D^{-1}CK^{-1}BD^{-1} \end{vmatrix} \quad (1.1.15)$$

if  $|D| \neq 0$ . In (1.1.14) and (1.1.15) the notation  $H = D - CA^{-1}B$  and  $K = A - BD^{-1}C$  is used.

**Theorem 1.1.5.** If  $B = A^{-1}$  then for any  $1 \leq i_1 < i_2 < \dots < i_p \leq n$ ,  $1 \leq k_1 < k_2 < \dots < k_p \leq n$ :

$$B \begin{pmatrix} i_1, i_2, \dots, i_p \\ k_1, k_2, \dots, k_p \end{pmatrix} = \frac{(-1)^{\sum_{v=1}^p i_v + \sum_{v=1}^p k_v} A \begin{pmatrix} k'_1, k'_2, \dots, k'_{n-p} \\ i'_1, i'_2, \dots, i'_{n-p} \end{pmatrix}}{A \begin{pmatrix} 1, 2, \dots, m \\ 1, 2, \dots, m \end{pmatrix}}, \quad (1.1.16)$$

where  $i_1 < i_2 < \dots < i_p$  together with  $i'_1 < i'_2 < \dots < i'_{n-p}$ , and  $k_1 < k_2 < \dots < k_p$  together with  $k'_1 < k'_2 < \dots < k'_{n-p}$  form a complete system of indices.

If  $B = A^{-1}$  and

$$B = \begin{vmatrix} B_{rr} & B_{rq} \\ B_{qr} & B_{qq} \end{vmatrix} \quad \text{and} \quad A = \begin{vmatrix} A_{rr} & A_{rq} \\ A_{qr} & A_{qq} \end{vmatrix},$$

then from Theorem 1.1.5 it follows, in particular, that

$$|B_{rr}| = |A_{qq}| / |A|. \quad (1.1.17)$$

#### IV. Positive-definite matrices.

**DEFINITION 11.** The square matrix  $A = \|A_{ik}\|_1^m$  is called symmetric if  $A_{ik} = A_{ki}$  (or  $A = A'$ ).

**DEFINITION 12.** A symmetric matrix  $A$  is called positive definite if the quadratic form  $Q(x) = x'Ax = \sum_{i,j=1}^m A_{ij}x_i x_j$  is positive for all nontrivial systems of values of real variables  $x_i$ , i.e., for

$$x' = \|x_1, x_2, \dots, x_m\| \neq \|0, 0, \dots, 0\|.$$

If under these conditions  $Q(x) \geq 0$ , then the matrix  $A$  is called positive semidefinite

In the remaining part of this section we will consider only symmetric matrices

**Theorem 1.1.6.** *In order that the matrix  $A$  of order  $n$  be positive definite it is necessary and sufficient that*

$$A \begin{pmatrix} 1, 2, \dots, k \\ 1, 2, \dots, k \end{pmatrix} > 0 \quad (k = 1, 2, \dots, n)$$

From this it is possible to verify that for positive-definite matrices all of its principal minors are greater than zero

$$A \begin{pmatrix} i_1, i_2, \dots, i_k \\ i_1, i_2, \dots, i_k \end{pmatrix} > 0 \quad (1 \leq i_1 < i_2 < \dots < i_k \leq n, \quad k = 1, 2, \dots, n) \quad (1.1.18)$$

In particular all of its diagonal elements must be greater than zero

**Theorem 1.1.7.** *In order that the matrix  $A$  be positive semidefinite, it is necessary and sufficient that all of its principal minors be nonnegative*

$$A \begin{pmatrix} i_1, i_2, \dots, i_k \\ i_1, i_2, \dots, i_k \end{pmatrix} \geq 0 \quad (1 \leq i_1 < i_2 < \dots < i_k \leq n, \quad k = 1, 2, \dots, n) \quad (1.1.19)$$

**DEFINITION 13** We consider the matrix  $\lambda I_m - A$ . The determinant of this matrix is called the characteristic polynomial of the matrix  $A = \|A_{ik}\|_1^m$ . The roots  $\lambda_1, \lambda_2, \dots, \lambda_m$  of the equation  $|\lambda I_m - A| = 0$  are called the characteristic numbers of the matrix  $A$ .

**Theorem 1.1.8.** *The matrix  $A = A'$  is positive definite (semidefinite), if and only if all of its characteristic numbers are positive (nonnegative), that is, when it can be represented in the form*

$$A = U \|\lambda_i \delta_{ik}\|_1^m U, \quad \lambda_i > 0 \quad (\lambda_i \geq 0), \quad (1.1.20)$$

where  $UU' = I_m$  ( $U$  is an orthogonal matrix)

From (1.1.20) it follows that the matrix  $A$  can be represented in the form

$$A = FF', \quad \text{where } F = U \|(\lambda_i)^{1/2} \delta_{ik}\|_1^m. \quad (1.1.21)$$

**Theorem 1.1.9.** *If the matrix  $A$  is represented in the form  $A = FF'$ , then it is positive semidefinite. If the matrix  $F$  is square and  $|F| > 0$ , then the matrix is positive definite.*

**Theorem 1.1.10.** *Let two matrices of the same order  $A$  and  $B$  be given, and let the matrix  $A$  be positive definite. Then there is a matrix  $T$  such that*

$$TBT' = A, \quad A_{ik} = \lambda_i \delta_{ik} \quad (i, k = 1, 2, \dots, m) \quad (1.1.22a)$$

and

$$TAT' = I. \quad (1.1.22b)$$

## V. Basic inequalities.

**Theorem 1.1.11 (Hadamard's Inequality).** *Let  $A$  be a real square matrix of order  $m$ ; then*

$$|A|^2 \leq \prod_{i=1}^m \sum_{j=1}^m A_{ij}^2. \quad (1.1.23)$$

Moreover, if  $A$  is positive definite, then

$$|A| \leq \prod_{i=1}^m A_{ii}, \quad (1.1.24)$$

and (generalized inequality of Hadamard)

$$|A| \leq A \begin{pmatrix} 1, 2, \dots, p \\ 1, 2, \dots, p \end{pmatrix} A \begin{pmatrix} p+1, \dots, n \\ p+1, \dots, n \end{pmatrix}. \quad (1.1.25)$$

In the future, the inequality  $C > 0$  ( $C \geq 0$ ) indicates that the matrix  $C$  is positive definite (semidefinite). The inequality  $A > B$  ( $A \geq B$ ) is equivalent to  $A - B > 0$  ( $\geq 0$ ).

**Theorem 1.1.12.** *If  $A$  and  $B$  are positive-definite matrices, then*

$$\alpha A^{-1} + (1 - \alpha) B^{-1} \geq [\alpha A + (1 - \alpha) B]^{-1}, \quad 0 < \alpha < 1. \quad (1.1.26)$$

Moreover, the equality sign holds only if  $A = B$ .

**Theorem 1.1.13.** If the matrix  $A_j$  has dimension  $n \times m$  and  $B_j$  is a square positive-definite matrix of order  $m$  ( $j = 1, 2$ ), then

$$\begin{aligned} & (1 - \alpha) A_1 B_1^{-1} A_1' + \alpha A_2 B_2^{-1} A_2' \\ & \geq [(1 - \alpha) A_1 + \alpha A_2] [(1 - \alpha) B_1 + \alpha B_2]^{-1} [(1 - \alpha) A_1' + \alpha A_2'], \\ & \quad 0 < \alpha < 1 \quad (1.1.27) \end{aligned}$$

**Theorem 1.1.14.** If  $A$  and  $B$  are positive-definite matrices of order  $m$  and  $0 < \alpha < 1$ , then

$$|\alpha A + (1 - \alpha) B| \geq |A|^{1-\alpha} |B|^{\alpha} \quad (1.1.28)$$

Moreover, equality holds only if  $A = B$

We indicate the sum of the diagonal elements of a square matrix  $A$  by  $\text{Tr } A$ .  $\text{Tr } A = \sum_{i=1}^m A_{ii}$ . It is easy to show that

$$\text{Tr } AB = \text{Tr } BA, \quad \text{Tr } UAU = \text{Tr } A, \quad (1.1.29)$$

where  $U$  is an orthogonal matrix. In particular, it follows from (1.1.29) that  $\text{Tr } A = \sum_{i=1}^m \lambda_i$ , where  $\lambda_i$  are the characteristic numbers of the matrix  $A$ .

**Theorem 1.1.15.** If  $A$  and  $B$  are positive-definite matrices of order  $m$ , then

$$\text{Tr } AB \geq m |A|^{1/m} |B|^{1/m} \quad (1.1.30)$$

**Theorem 1.1.16.** If  $A > B$ , then  $|A| > |B|$

**Theorem 1.1.17.** If  $A$  and  $B$  are positive-semidefinite matrices of rank  $r_A$  and  $r_B$ , respectively, then the matrix  $C = A + B$  has rank  $r_C \leq r_A + r_B$ .

## VI. Differentiation and integration of matrices

**DEFINITION 14** The elements of the matrix  $\partial A / \partial t$  are equal to the derivatives of the corresponding elements of the matrix  $A$

$$\left( \frac{\partial A}{\partial t} \right)_{ik} = \left( \frac{\partial A_{ik}}{\partial t} \right) \quad (i = 1, 2, \dots, m, k = 1, 2, \dots, n)$$

It immediately follows from the definition that

$$(\partial/\partial t)(AB) = (\partial A/\partial t)B + A(\partial B/\partial t), \quad (1.1.31)$$

$$(\partial/\partial t) \operatorname{Tr} A = \operatorname{Tr}(\partial/\partial t)A. \quad (1.1.32)$$

If  $A$  is a square matrix and  $|A| \neq 0$ , then

$$(\partial/\partial t)A^{-1} = -A^{-1}(\partial A/\partial t)A^{-1}, \quad (1.1.33)$$

$$(\partial/\partial t)\ln|A| = \operatorname{Tr} A^{-1}(\partial A/\partial t). \quad (1.1.34)$$

**DEFINITION 15.** Elements of the matrix  $\int A dt$  are equal to the integrals of the corresponding elements of the matrix  $A$ :  $(\int A dt)_{ik} = \int A_{ik} dt$  ( $i = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, n$ ).

## 1.2. General Requirements Satisfied by Estimates

We will assume that the results of measurements are random variables such that

$$E(y|x) = \eta(x, \theta), \quad (1.2.1)$$

where  $y$  is the result of a measurement at the point  $x$ ,  $\eta(x, \theta)$  is a function, the analytic form of which is known,  $\theta' = \|\theta_1, \dots, \theta_m\|$  are unknown parameters, and the operator  $E$  corresponds to the expectation operator.

Let the goal of the given experimental analysis be the determination of an estimate of the unknown parameter  $\theta$  or estimates of the response surface  $\eta(x, \theta)$  in some given region  $X_0$ . Experiments subject to the assumption (1.2.1) will be called regression experiments if their goal is to find estimates of unknown parameters or an unknown response surface. The procedure of finding these estimates will be called regression analysis.

By an estimate of some unknown quantity  $\tau$  we will understand a quantity  $\tilde{\tau}$  which is close to the true value  $\tau_0$ . The concept of "closeness" will be made precise later.

It is evident that  $\tilde{\theta}$  [or  $\tilde{\eta}(x, \theta)$ ] must in some way depend on the results of the measurements, i.e.,

$$\tilde{\theta} = \Psi(y_1|x_1, y_2|x_2, \dots, y_n|x_n), \quad (1.2.2)$$

where  $y_i$  is the measurement taken at the point  $x_i$ . Some values  $x_i$

can coincide for distinct  $i$ . Henceforth, the collection  $\{y_1, y_2, \dots, y_n\}$  will be called the sample  $Y_n$ .

We must seek a functional  $\Psi$  possessing a series of given properties which would give the estimates. Such properties are as follows (cf., for example, [I-II])

*1 Consistency* The estimate  $\hat{\theta}$  is called consistent if, for any two given positive numbers  $\gamma$  and  $\beta$  ( $\beta < 1$ ), there is a sufficiently large sample size (we note that by sample size is understood the number of elements in the collection  $\{y_1, y_2, \dots, y_n\}$ ), such that

$$P(|\hat{\theta}_n - \theta_{n0}| < \gamma) > 1 - \beta, \quad (1.2.3)$$

where  $\theta_{n0}$  is the true value of the  $n$ th parameter. A short definition of consistency can be formulated as follows

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta_{n0}| \leq \epsilon) = 1 \quad (1.2.4)$$

*2 Unbiasedness* An estimate is called unbiased (more precisely, unbiased in the mean) if

$$E_{r_n}(\hat{\theta}) = \theta_0 \quad (1.2.5)$$

*3 Sufficiency* An estimate  $\hat{\theta}$  is called sufficient if the conditional distribution  $p(\hat{\theta} | \hat{\theta}_0)$ , where  $\hat{\theta}$  is any other estimator, does not depend on  $\theta_0$ . Roughly speaking the definition of sufficiency requires that the estimate contain all of the information about the sought parameters included in the results of the observations.

*4 Efficiency* An unbiased estimate  $\hat{\theta}$  is called efficient if, for the dispersion matrix of this estimate

$$D(\hat{\theta}) = E_{r_n}[(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T] \quad (1.2.6)$$

the inequality

$$D(\hat{\theta}) - D(\tilde{\theta}) \geq 0 \quad (1.2.7)$$

holds, where  $D(\tilde{\theta})$  is the dispersion matrix of any other estimator  $\tilde{\theta}$ .

It is evident that for any  $\eta(x, \theta)$  and each density function  $p(y | x)$ , there will be, generally speaking, a "best" estimate, i.e., an estimate for which all (or at least most) of Properties 1-4 are satisfied. Such dependence of the best estimator on the form of  $\eta(x, \theta)$  and  $p(y | x)$  is very inconvenient in practice (each experimental situation has its own method of analysis and experimental design). Therefore, it is

reasonable to waive satisfaction of some of Properties 1–4, and to construct estimates which would not be sensitive to the form of  $\eta(x, \theta)$  and  $p(y | x)$ . Independence of the form of  $p(y | x)$  is usually important, since, in practice, the function  $p(y | x)$  is usually unknown.

### 1.3. Best Linear Estimates

We will assume that the function  $\eta(x, \theta)$  is linear in the parameters, that is,

$$E(y | x) = \eta(x, \theta) = \theta' f(x), \quad (1.3.1)$$

where  $f'(x) = \|f_1(x), f_2(x), \dots, f_m(x)\|$  are known functions. We will also assume that at the points  $x_1, x_2, \dots, x_n$ , independent measurements  $y_1, y_2, \dots, y_n$  were taken with dispersions  $b_1^2, b_2^2, \dots, b_n^2$ .

We emphasize that no assumption on the form of  $p(y | x)$  is made, besides the existence of finite second moments. We restrict our considerations only to linear estimates for  $\theta_0$ , i.e., to estimates which can be represented in the form

$$\hat{\theta} = Ty, \quad (1.3.2)$$

where  $y' = \|y_1, y_2, \dots, y_n\|$ . It can be shown that it is possible to find an estimator  $\hat{\theta}$  for  $\theta_0$ , which is robust, consistent, unbiased, and has minimum dispersion  $D(\hat{\theta}_a)$  among the collection of all linear unbiased estimators. This estimate is called the *best linear estimate*.

Before we go on to a presentation of the basic theorems on best linear estimators we formulate the following well-known theorem in probability theory.

**Theorem 1.3.1.** *Let the random variable  $u$  be a linear combination of random variables  $v$ :*

$$u = Lv. \quad (1.3.3)$$

*Then*

(1) *the mean values  $E(u)$  and  $E(v)$  are related to one another by the relation*

$$E(u) = LE(v); \quad (1.3.4)$$

(2) *the dispersion matrices*

$$D(u) = E\{[u - E(u)][u - E(u)]'\},$$

and

$$D(v) = E[(v - E(v))(v - E(v))]$$

are related by the relationship

$$D(u) = LD(v)L' \quad (1.3.5)$$

The proof of Theorem 1.3.1 can be found, for example, in [II]. We now prove the following theorem, basic to best linear unbiased estimators [9].

**Theorem 1.3.2.** Under the above assumptions [cf. formula (1.3.1) and the explanation following it] the best linear unbiased estimator for the unknown parameter  $\theta$  is

$$\hat{\theta} = M^{-1}Y, \quad (1.3.6)$$

where the matrix  $M$  is assumed to be nondegenerate and equals

$$M = \sum_{i=1}^n w_i f(x_i) f'(x_i) \quad (1.3.7)$$

$$Y = \sum_{i=1}^n w_i y_i f(x_i), \quad (1.3.8)$$

and  $w_i = b_i^{-2}$

The dispersion matrix of the estimator  $\hat{\theta}$  is equal to

$$D(\hat{\theta}) = M^{-1} \quad (1.3.9)$$

*Proof* We show that the estimate (1.3.6) is unbiased. By Theorem 1.3.1

$$E(\hat{\theta}) = E(M^{-1}Y) = M^{-1}E(Y) \quad (1.3.10)$$

Using (1.3.1) and the manifest form of  $Y$ , we obtain

$$\begin{aligned} E(Y) &= E\left[\sum_{i=1}^n w_i f(x_i) y_i\right] = \sum_{i=1}^n w_i f(x_i) E(y_i) \\ &= \sum_{i=1}^n w_i f(x_i) f'(x_i) \theta_0 = M\theta_0 \end{aligned} \quad (1.3.11)$$

We set (1.3.11) into (1.3.10):  $E(\tilde{\theta}) = M^{-1}M\theta_0 = \theta_0$ , i.e., the estimate  $\tilde{\theta}$  is unbiased.

We will now prove that the dispersion  $D(\theta_\alpha)$  is smallest among all linear unbiased estimators. We consider any arbitrary linear unbiased estimator

$$\tilde{\theta} = Ty. \quad (1.3.12)$$

Since the estimator  $\tilde{\theta}$  is unbiased

$$E(\tilde{\theta}) = E(Ty) = \theta_0, \quad (1.3.13)$$

or, using (1.3.1),

$$E(Ty) = TE(y) = TF'\theta_0 = \theta_0, \quad (1.3.14)$$

where

$$F = \|f(x_1), f(x_2), \dots, f(x_n)\|. \quad (1.3.15)$$

From (1.3.14) the auxiliary equality, which will be necessary later, follows:

$$TF' = I, \quad (1.3.16)$$

where  $I$  is the identity matrix.

By Theorem 1.3.1 and Eq. (1.3.12) the dispersion matrix

$$D(\tilde{\theta}) = TD(y) T', \quad (1.3.17)$$

where

$$D(y) = \begin{vmatrix} b_1^2 & 0 & \cdots & 0 \\ 0 & b_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n^2 \end{vmatrix}.$$

We introduce the matrices

$$\phi = F\Sigma^{-1} \quad \text{and} \quad \mathcal{T} = T\Sigma, \quad (1.3.18)$$

where

$$\Sigma = \begin{vmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{vmatrix}, \quad \Sigma\Sigma' = D(y).$$

Equations (1.3.16) and (1.3.17) can be rewritten in the forms

$$\mathcal{F}\phi = I \quad (1.3.19)$$

and

$$D(\theta) = \mathcal{F}\mathcal{F} \quad (1.3.20)$$

The following identity is valid

$$D(\theta) = \mathcal{F}\mathcal{F} = \{(\phi\phi)^{-1}\phi\}\{(\phi\phi)^{-1}\phi\} + \{\mathcal{F} - (\phi\phi)^{-1}\phi\}\{\mathcal{F} - (\phi\phi)^{-1}\phi\} \quad (1.3.21)$$

The identity (1.3.21) is easy to verify, removing braces and making use of (1.3.19).

$$\begin{aligned} & (\phi\phi)^{-1}\phi\phi(\phi\phi)^{-1} + \mathcal{F}\mathcal{F} - (\phi\phi)^{-1}\phi\mathcal{F} \\ & - \mathcal{F}\phi(\phi\phi)^{-1} + (\phi\phi)^{-1}\phi\phi(\phi\phi)^{-1} \\ & = (\phi\phi)^{-1} + \mathcal{F}\mathcal{F} - (\phi\phi)^{-1}I - I(\phi\phi)^{-1} + (\phi\phi)^{-1} = \mathcal{F}\mathcal{F} \end{aligned}$$

Each of the terms on the right hand side of (1.3.21) is a matrix of the type  $AA$ , the diagonal elements of which,

$$\{AA\}_{aa} = \sum_b A_{ab}A_{ba} = \sum_b A_{aa}^2,$$

are not less than 0. It follows that  $D_{aa}(\theta)$  will have a smallest value when

$$\mathcal{F}_0 = (\phi\phi)^{-1}\phi \quad (1.3.22)$$

From (1.3.22) it follows that

$$T_0 = [FD^{-1}(y)F]^{-1}FD^{-1}(y) \quad (1.3.23)$$

or, using (1.3.12) and (1.3.15),

$$\theta = T_0y = [FD^{-1}(y)F]^{-1}FD^{-1}(y)y = M^{-1}Y \quad (1.3.24)$$

where  $M$  and  $Y$  are defined by Eqs (1.3.7) and (1.3.8). From (1.3.17) and (1.3.24) it follows that

$$D(\theta) = [FD^{-1}(y)F]^{-1} = M^{-1} \quad (1.3.25)$$

Comparing formulas (1.3.24) and (1.3.25) with formulas (1.3.6)–(1.3.8), one may easily obtain the validity of the theorem.

The matrix

$$M = \sum_{i=1}^n w_i f(x_i) f'(x_i) = \sum_{i=1}^n M(x_i) \quad (1.3.26)$$

is called the Fisher information matrix. If the matrix  $M$  is not degenerate, then its inverse matrix is the dispersion matrix of the best linear estimator  $\hat{\theta}$ . In constructing the mathematical apparatus of designing experiments, this matrix will be met very often. We point out the following important property of the information matrix.

**Corollary 1.** *The Fisher information matrix is positive semidefinite.*

Indeed  $M = \sum_{i=1}^n w_i f(x_i) f'(x_i) = \phi\phi'$ . But by Theorem 1.1.9, any matrix which can be represented in the form  $AA'$  is positive semi-definite.

The estimated parameter  $\theta$  is a vector variable and, generally speaking, the accuracy of the estimate  $\hat{\theta}$  is characterized by all elements of the dispersion matrix  $D(\hat{\theta})$ . Therefore, distinct estimates for  $\theta$  can be compared not only according to the diagonal elements  $D_{\alpha\alpha}(\hat{\theta})$ , as was done above, but also by other methods. We introduce two of the most widespread methods of comparing estimators.

(1) The estimator  $\tilde{t}$  is preferred to the estimator  $\tilde{t}'$  if

$$D(\tilde{t}') = D(\tilde{t}) + D, \quad (1.3.27)$$

where  $D$  is some positive-definite matrix, or abbreviated

$$D(\tilde{t}') > D(\tilde{t}). \quad (1.3.28)$$

(2) The estimator  $\tilde{t}$  is preferred to the estimator  $\tilde{t}'$  if

$$|D(\tilde{t}')| > |D(\tilde{t})|. \quad (1.3.29)$$

The determinant  $|D(\tilde{t})|$  is called the generalized dispersion of the estimator  $\tilde{t}$ .

From the proof of Theorem 1.3.2 it immediately follows that the best linear estimator is optimal in the sense of (1.3.27) and (1.3.29). More precisely, these facts can be formulated in the following way.

**Corollary 2.** *The best linear unbiased estimator (1.3.6) has the*

smallest dispersion matrix (cf. Section 1.1, Part V) among all linear unbiased estimators  $\hat{\theta}$ , i.e.,

$$D(\hat{\theta}) - D(\tilde{\theta}) \leq 0 \quad (1.3.30)$$

In other words, the best unbiased estimator is efficient among the class of unbiased estimators. Formula (1.3.30) is an obvious corollary of formula (1.3.21).

**Corollary 3** *The determinant of the dispersion matrix of the best linear estimate (1.3.6) is smallest for all linear unbiased estimators*

$$|D(\hat{\theta})| < |D(\tilde{\theta})| \quad (1.3.31)$$

The result (1.3.31) follows immediately from (1.3.30) and an application of Theorem 1.1.16. We will give two further useful corollaries of Theorem 1.3.2.

**Corollary 4** *The best linear unbiased estimate of an arbitrary linear combination  $t = C\theta$  is  $\hat{t} = C\hat{\theta}$ . The dispersion matrix of the estimator  $\hat{t}$  is equal to*

$$D(\hat{t}) = CD(\hat{\theta})C \quad (1.3.32)$$

*If  $t$  is any linear unbiased estimate of the parameters  $\theta$  distinct from  $\hat{t}$ , then*

- (1)  $D_{ss}(t) \geq D_{ss}(\hat{t})$
- (2)  $D(t) - D(\hat{t}) \leq 0,$  (1.3.33)
- (3)  $|D(t)| \leq |D(\hat{t})|$

The proof of Corollary 4 is easy to carry out by making use of Theorem 1.3.1 and formula (1.3.21). Indeed, applying the formula for transformation of dispersions (1.3.5) to both parts of (1.3.21) and then repeating word for word the argument used in Theorem 1.3.2 and its Corollaries 2 and 3, we arrive at the results (1.3.32) and (1.3.33).

*Remark* In some cases, an estimator  $\hat{t}$  can be found when the information matrix is singular and a direct application of the formula  $\hat{t} = C\hat{\theta}$  is not possible.

There exist several methods for computing these estimators and their dispersion matrices [13, 31]. In the future, the method (more precisely, its generalization) presented in [31] will be useful for us.

Let  $\tilde{M}$  be any positive-definite matrix.

Then

$$\hat{\tau} = \lim_{\alpha \rightarrow 0} C[M + \alpha \tilde{M}]^{-1} Y,$$

and

$$D(\hat{\tau}) = \lim_{\alpha \rightarrow 0} C[M + \alpha \tilde{M}]^{-1} C.$$

It may be verified that the corresponding limits do not depend on the choice of  $\tilde{M}$ .

**Corollary 5.** *The best linear estimate of the function  $\eta(x, \theta)$  at an arbitrary point  $x$  is*

$$\hat{\eta}(x, \theta) = f'(x) \hat{\theta}. \quad (1.3.34)$$

*The dispersion of the estimate  $\hat{\eta}(x, \theta)$  equals*

$$d(x) = f'(x) D(\hat{\theta}) f(x). \quad (1.3.35)$$

Essentially, Corollary 5 is an important particular case of Corollary 4 with  $C = f'(x)$ .

In the future, we shall denote the quantity  $[d(x)]^{1/2}$  by  $b(x)$  and call it the standard deviation. Very often in experimental practice, at each point  $x_i$  several independent measurements  $y_{i1}, y_{i2}, \dots, y_{in_i}$  are taken, each with dispersion  $b_i^2$ . It can be shown that for constructing best linear estimates the experimenter need not keep all values  $y_{ij}$  ( $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, n_i$ ). It is sufficient to keep only the value

$$y_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}. \quad (1.3.36)$$

In order to show this we will prove the following assertion.

**Corollary 6.** *If at the point  $x_i$  ( $i = 1, 2, \dots, n$ ) the results  $y_{i1}, y_{i2}, \dots, y_{in_i}$  are obtained, the best linear estimate for  $\theta$  will be*

$$\hat{\theta} = M^{-1} Y,$$

where the matrix  $M$  is assumed to be nondegenerate and equal to

$$\begin{aligned} M &= \sum_{i=1}^n w_i f(x_i) f'(x_i), \\ Y &= \sum_{i=1}^n w_i y_i f(x_i), \end{aligned} \quad (1.3.37)$$

where  $w_i = n_i/b_i^2$ , and  $y_i$  is defined by formula (1.3.36)

Formulas (1.3.7) and (1.3.8) can in this case be written in the form

$$M = \sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} f(x_i) f'(x_i)$$

or, using (1.3.37),

$$M = \sum_{i=1}^n n_i b_i^{-2} f(x_i) f'(x_i) = \sum_{i=1}^n w_i f(x_i) f'(x_i) \quad (1.3.38)$$

and

$$\begin{aligned} Y &= \sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} y_{ij} f(x_i) = \sum_{i=1}^n n_i b_i^{-2} \sum_{j=1}^{n_i} n_i^{-1} y_{ij} f(x_i) \\ &= \sum_{i=1}^n w_i y_i f(x_i) \end{aligned} \quad (1.3.39)$$

By comparing (1.3.38) and (1.3.39) with (1.3.37) the validity of our assertion is shown

Formula (1.3.37) is especially useful for large-dimensional experimental data, since it alleviates the necessity of introducing essentially unnecessary information into the electronic computing machine's memory

We note one further important property of best linear estimates

**Theorem 1.3.3** *The best linear estimate  $\hat{\theta}$  minimizes the weighted sum of the squares of the deviations*

$$S(\hat{\theta}) = \sum_{i=1}^n w_i [y_i - f(x_i) \hat{\theta}]^2 \quad (1.3.40)$$

*Proof.* We find the minimum of  $S(\theta)$  in  $\theta$ . Differentiating both sides of (1.3.40) with respect to  $\theta_\alpha$  ( $\alpha = 1, 2, \dots, m$ ) and setting the partial derivatives equal to zero, we obtain

$$\sum_{i=1}^n w_i f_\alpha(x_i) \left[ y_i - \sum_{\beta=1}^m f_\beta(x_i) \theta_\beta \right] = 0,$$

or, in matrix form,

$$\sum_{i=1}^n w_i f(x_i) [y_i - f'(x_i) \theta] = 0.$$

Removing parentheses, we obtain

$$\sum_{i=1}^n w_i y_i f(x_i) = \sum_{i=1}^n w_i f(x_i) f'(x_i) \theta.$$

Using the notation (1.3.7) and (1.3.8) we obtain

$$Y = M\theta \quad \text{or} \quad \hat{\theta} = M^{-1}Y,$$

which proves the theorem.

Theorem 1.3.3 can be formulated somewhat differently, indeed: In the case of a linear parametrization, the best linear estimate  $\hat{\theta}$  coincides with the estimate obtained according to the method of least squares. Theorem 1.3.3 and Corollary 4 show that best linear estimates, which are "best" in the space of parameters  $\hat{\theta}$ , simultaneously turn out to be "best" also in the space of control variables (i.e., they minimize  $d(x)$  and  $S(\theta) = \sum_{i=1}^n w_i [y_i - f'(x_i)\theta]^2$ ). This property of best linear estimates will be extensively used in constructing the mathematical apparatus of designing experiments.

It must be pointed out that, in most cases, best linear estimates are not sufficient. For some distributions  $p(y | x)$ , for example, for the normal law, they are sufficient.

It can be shown [62] that the best linear estimate is consistent if

$$\lim_{N \rightarrow \infty} N^{-1}M = M_f \quad \text{and} \quad M_f \neq 0,$$

where

$$N = \sum_{i=1}^n n_i$$

**EXAMPLE** Let  $E(y|x) = \theta_1 + \theta_2x$  and let measurements be taken at the points  $x_1 = -1$ ,  $x_2 = 0$ ,  $x_3 = +1$  with dispersions  $b_1^2 = 8$ ,  $b_2^2 = 1$ ,  $b_3^2 = 8$

We consider two estimates for the parameters  $\theta_\alpha$  ( $\alpha = 1, 2$ ). For the first, we choose the best linear unbiased estimator. For this estimator, the dispersion matrix, corresponding to Theorem 1.3.2, equals

$$D(\hat{\theta}) = \left[ \sum_{i=1}^3 w_i f(x_i) f'(x_i) \right]^{-1} = \begin{vmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{vmatrix}^{-1} = \begin{vmatrix} 1 & 0 \\ 0 & 4 \end{vmatrix},$$

where  $f'(x) = \|1, x\|$

As the second estimator for  $\theta$ , we choose the value  $\bar{\theta}$ , which minimizes the quadratic form

$$S(\theta) = \sum_{i=1}^3 [y_i - (\theta_1 + \theta_2 x_i)]^2$$

Sometimes these estimates are called estimates constructed according to the method of least squares without calculating weights. Differentiating  $S(\theta)$  with respect to  $\theta_1$  and  $\theta_2$ , after a simple calculation we obtain  $\bar{\theta} = Ty$ , where

$$T = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{vmatrix}$$

It is easy to show that the estimate  $\bar{\theta}$  is unbiased. Indeed

$$F = \|f(x_1) f(x_2) f(x_3)\| = \begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{vmatrix}$$

and it follows that  $TF = I_3$ , which corresponds to the condition of unbiasedness (1.3.16). From (1.3.17) it is possible to define the dispersion matrix of the estimator  $\bar{\theta}$

$$D(\bar{\theta}) = TD(y) T = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{vmatrix} \begin{vmatrix} 8 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 8 \end{vmatrix} \begin{vmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & 0 \\ 0 & 4 \end{vmatrix}$$

In this way  $D(\bar{\theta}) \geq D(\hat{\theta})$

We compare the characteristics of the estimates  $\hat{\theta}$  and  $\bar{\theta}$  in the space of measurements. According to Theorem 1.3.1 the dispersion of

the estimated line  $\theta_1 + \theta_2 x$  at the point  $x$  equals  $d_1(x) = 1 + 4x^2$  for the estimate  $\hat{\theta}$  and  $d_2(x) = \frac{5}{2}x^2 + 4x^2$  for the estimate  $\tilde{\theta}$ . Therefore, for any  $x$ ,  $d_1(x) < d_2(x)$ .

#### 1.4. The Search for Estimates for Nonlinear Parametrization. Best Quasi-Linear Estimates

Earlier, a method was described in detail for finding estimates for the case of linear parametrization when condition (1.3.1) was valid. In many complicated physical, chemical, or other experiments the function  $\eta(x, \theta)$  is nonlinear in some unknown parameters. The problem of seeking estimates in this case becomes significantly more complicated.

The most widespread method of finding estimates for nonlinear parametrization is the method of maximum likelihood (a detailed exposition of it can be found, for example, in [5, 9, 10]). This method gives consistent, asymptotically normal and efficient estimates under very general assumptions. However, it depends on knowledge of the density function  $p(y | x)$ . As previously noted (cf. Sections 1.2 and 1.3), this requirement is a strong limitation. Therefore, we will turn to somewhat coarser methods than the method of maximum likelihood, but which will depend only on the values of the dispersion.

Let  $\eta(x, \theta)$  be nonlinear in  $\theta$ . We will assume that the situation is the same as presented in the remarks following formula (1.2.1); all measurements belonging to  $Y_n$  are independent, and all measurements taken at the same point  $x_i$  have the same dispersion  $b_i^2$ . It is known that the nondegenerate solution  $\hat{\theta}$  of the equation

$$\sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} [y_{ij} - \eta(x_i, \hat{\theta})]^2 = \min_{\theta} \sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} [y_{ij} - \eta(x_i, \theta)]^2 \quad (1.4.1)$$

is a consistent estimator for  $\theta_0$ , i.e., for sufficiently large sample size  $Y_n$ , the value  $\hat{\theta}$  is close to  $\theta_0$ .

It is easy to show that

$$\sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} [y_{ij} - \eta(x_i, \theta)]^2 = \sum_{i=1}^n w_i [y_i - \eta(x_i, \theta)]^2 + \sum_{i=1}^n \sum_{j=1}^{n_i} b_i^{-2} [y_{ij} - y_i]^2, \quad (1.4.2)$$

where  $y_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$  and  $w_i = n_i b_i^{-2}$ . From (1.4.2) it follows that

the solution of Eq. (1.4.1) coincides with the solution of the equation

$$\sum_{i=1}^n w [y_i - \eta(x_i, \theta)]^2 = \min_{\theta} \sum_{i=1}^n w [y_i - \eta(x_i, \theta)]^2 \quad (1.4.3)$$

Let the function  $\eta(x_i, \theta)$  be smooth in  $\theta$  in a neighborhood of the true value of the parameter  $\theta_0$  and let equation (1.4.3) have a unique solution (the case of the existence of several solutions will be considered in Chapter 7) then (1.4.3) is equivalent to the equation

$$\nabla \left\{ \sum_{i=1}^n w [y_i - \eta(x_i, \theta)]^2 \right\} = 0 \quad (1.4.4)$$

where the operator

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \frac{\partial}{\partial \theta_2} \\ \vdots \\ \frac{\partial}{\partial \theta_m} \end{bmatrix}$$

Let  $\hat{\theta}$  be the solution of this equation and let the sample size  $Y_n$  be such that  $\hat{\theta}$  belongs to the region where

$$\begin{aligned} \eta(x, \theta) &\simeq \eta(x, \theta_0) + (\theta - \theta_0) f(x) \\ f(x) &= \nabla \eta(x, \theta_0) \end{aligned} \quad (1.4.5)$$

Then from (1.4.4) it follows that

$$\sum_{i=1}^n w_i [y_i - \eta(x_i, \theta_0)] f(x_i) = \sum_{i=1}^n w_i f(x_i) f'(x_i) (\theta - \theta_0) \quad (1.4.6)$$

or

$$\theta - \theta_0 = M^{-1} Y \quad (1.4.7)$$

where

$$M = \sum_{i=1}^n w_i f(x_i) f'(x_i)$$

$$Y = \sum_{i=1}^n w [y_i - \eta(x_i, \theta_0)] f(x_i) \quad (1.4.8)$$

$$D(\hat{\theta}) \simeq M^{-1}$$

The inequality

$$\sum_{i=1}^n w_i \frac{\partial \eta(x_i, \theta)}{\partial \theta_\alpha} \frac{\partial \eta(x_i, \theta)}{\partial \theta_\beta} \Big|_{\theta=\theta_0} \gg \sum_{i=1}^n w_i^{1/2} \frac{\partial^2 \eta(x_i, \theta)}{\partial \theta_\alpha \partial \theta_\beta} \Big|_{\theta=\theta_0} \quad (1.4.9)$$

can serve as a simple criterion for changing (1.4.4) into (1.4.7).

Comparing (1.4.7) and (1.4.8) with (1.3.8) and (1.3.9) it is not difficult to see that in the case of nonlinear parametrization the estimates obtained by the method of least squares asymptotically (for sufficiently large sample size  $Y_n$ ) satisfy the criteria of optimality that best linear estimates satisfy.

One essential distinction between the case of linear and nonlinear parametrization should be pointed out. For linear parametrization, the dispersion matrix of the estimates [cf. (1.3.9)] *does not depend* on the value  $\theta_0$ . For nonlinear parametrization the dispersion matrix of the estimator *depends* [cf. (1.4.8) and (1.4.5)] on the value  $\theta_0$ . For this reason, formula (1.4.8) cannot be immediately applied to the computation of the dispersion matrix  $D(\hat{\theta}) = M^{-1}$ . We can, however, make use of the fact that the function  $\eta(x, \theta)$  is smooth near  $\theta_0$ , and therefore

$$f(x, \hat{\theta}) \rightarrow f(x, \theta_0) \quad \text{for } \hat{\theta} \rightarrow \theta_0. \quad (1.4.10)$$

From (1.4.8) and (1.4.10) it follows that

$$M(\theta_0) \simeq M(\hat{\theta}) = \sum_{i=1}^n w_i f(x_i, \hat{\theta}) f'(x_i, \hat{\theta})$$

or

$$D(\hat{\theta}) \simeq M^{-1}(\hat{\theta}). \quad (1.4.11)$$

An analogous transformation of  $\theta_0$  to  $\hat{\theta}$  is necessary for the use of inequality (1.4.9).

If the function  $\eta(x, \theta)$  is smooth in a closed region  $\Omega$ , containing the point  $\theta_0$ , then, using any point in this region as an initial approximation for  $\hat{\theta}$ , it is possible to construct an iterative process for solving equation (1.4.4), the basis of which is the formula closely related to (1.4.7):

$$\theta(s) = \theta(s-1) + M^{-1}[\theta(s-1)] Y[\theta(s-1)]. \quad (1.4.12)$$

Here

$$\begin{aligned} M[\theta(s)] &= \sum_{i=1}^n w_i f[x_i, \theta(s)] f[x_i, \theta(s)], \\ Y[\theta(s)] &= \sum_{i=1}^n w_i y_i f[x_i, \theta(s)] \end{aligned} \quad (14.13)$$

It is possible to show that (cf [14, 15])

$$\lim_{s \rightarrow \infty} \theta(s) = \theta$$

### 1.5. Estimation of the Dispersion of the Resulting Observations. The Efficiency of an Experiment

The results of Sections 1.2–1.4 are obtained under the assumption that the dispersion of each of the observations is given. In this section we generalize these results to the case of unknown dispersions. As above, let  $y_1, y_2, \dots, y_n$  be the results of observations at the points  $x_1, x_2, \dots, x_n$ .

We consider the case of equal dispersions  $b_i^2 = b^2$  ( $i = 1, 2, \dots, n$ ). The following theorem holds.

**Theorem 1.5.1.** If

$$E(y|x) = \theta f(x), \quad (15.1)$$

and

$$D(y|x) = b^2, \quad (15.2)$$

then

(1) the best linear estimate for  $\theta$  will be

$$\theta = M^{-1}Y, \quad (15.3)$$

where

$$M = \sum_{i=1}^n f(x_i) f(x_i), \quad (15.4)$$

$$Y = \sum_{i=1}^n y_i f(x_i)$$

(2) an unbiased estimate for the dispersion  $b^2$  will be

$$\hat{b}^2 = (n - m)^{-1} \sum_{i=1}^n [y_i - \hat{\theta}' f(x_i)]^2. \quad (1.5.5)$$

*Proof.* From Theorem 1.3.2 it follows that the best linear estimate is equal to

$$\begin{aligned} \hat{\theta} &= \left[ \sum_{i=1}^n w_i f(x_i) f'(x_i) \right]^{-1} \left[ \sum_{i=1}^n w_i y_i f(x_i) \right] \\ &= w^{-1} \left[ \sum_{i=1}^n f(x_i) f'(x_i) \right]^{-1} w \left[ \sum_{i=1}^n y_i f(x_i) \right] = M^{-1} Y, \end{aligned} \quad (1.5.6)$$

where  $w_i = w = b^{-2}$ .

In this manner, in the case of measurements with equal dispersions the estimate  $\hat{\theta}$  does not depend on the value of  $b^2$ .

We now prove that the estimator  $\hat{b}^2$  is unbiased.

$$\begin{aligned} E(\hat{b}^2) &= (n - m)^{-1} E \left\{ \sum_{i=1}^n [y_i - \hat{\theta}' f(x_i)]^2 \right\} \\ &= (n - m)^{-1} E \left\{ \sum_{i=1}^n [y_i - \theta_0' f(x_i) + \theta_0' f(x_i) - \hat{\theta}' f(x_i)]^2 \right\} \\ &= (n - m)^{-1} E \left\{ \sum_{i=1}^n [y_i - \theta_0' f(x_i)]^2 \right\} \\ &\quad + (n - m)^{-1} E \left\{ \sum_{i=1}^n [\theta_0' f(x_i) - \hat{\theta}' f(x_i)]^2 \right\} \\ &\quad - \frac{2}{n - m} E \sum_{i=1}^n [y_i - \theta_0' f(x_i)][(\hat{\theta} - \theta_0)' f(x_i)]. \end{aligned} \quad (1.5.7)$$

The first term can be computed using (1.5.2):

$$E \left\{ \sum_{i=1}^n [y_i - \theta_0' f(x_i)]^2 \right\} = \sum_{i=1}^n E\{[y_i - \theta_0' f(x_i)]^2\} = nb^2. \quad (1.5.8)$$

We compute the second term

$$E \left\{ \sum_{i=1}^n [\theta_0 f(x_i) - \hat{\theta} f(x_i)]^2 \right\} = F \left\{ \sum_{i=1}^n f(x_i)(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta}) f(x_i) \right\} \quad (159)$$

By (1129) the expression (159) can be rewritten in the form

$$\begin{aligned} & F \left\{ \sum_{i=1}^n \text{Tr}[f(x_i) f(x_i) (\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})] \right\} \\ & = \text{Tr} \left\{ \sum_{i=1}^n f(x_i) f(x_i) E[(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})] \right\} \quad (1510) \end{aligned}$$

Since (cf Theorem 132)

$$D^{-1}(\hat{\theta}) = w \sum_{i=1}^n f(x_i) f(x_i),$$

and by definition

$$E[(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})] = D(\hat{\theta})$$

it follows that

$$\text{Tr} \left\{ \sum_{i=1}^n f(x_i) f(x_i) E[(\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})] \right\} = \text{Tr} b^2 I_m = mb^2, \quad (1511)$$

where  $m$  is the number of unknown parameters. The mixed terms give

$$\begin{aligned} & E \sum_i [y_i - \theta_0 f(x_i)](\hat{\theta} - \theta_0) f(x_i) \\ & = E \sum_i y_i (\hat{\theta} - \theta_0) f(x_i) \\ & = E(\hat{\theta} - \theta_0) M \hat{\theta} \\ & = E(\hat{\theta} - \theta_0) M (\hat{\theta} - \theta) \\ & = mb^2 \end{aligned}$$

Combining this last result with (157), (158) and (1511), we obtain  $E b^2 = b^2$ , i.e., the estimate  $\hat{b}^2$  is unbiased.

If the region  $X$ , in which the observations are taken, is comparatively small and the experimental conditions inside it are almost constant,

then the assumption of equal dispersions of the measurements is valid. In describing the mechanism of a phenomenon, experimenters, as a rule, are interested in a broad region  $X$  in which the assumption about homogeneity of experimental conditions is not fulfilled. However, in many cases it is possible to construct a function called *the efficiency of an experiment*,  $\lambda(x)$ , which permits the comparison of the values of the errors (dispersions) of the measurements at various points:

$$\tilde{\lambda}(x) = b^{-2}\lambda(x) = b^{-2}(x). \quad (1.5.12)$$

In those cases when the constant  $b^2$  is unknown it is possible to find an estimate of it.

**Theorem 1.5.2.** *If*

$$E(y | x) = \theta' f(x) \quad \text{and} \quad D(y | x) = b^2 \lambda^{-1}(x), \quad (1.5.13)$$

*then*

(1) *the best linear estimate for  $\theta$  will be*

$$\hat{\theta} = M^{-1}Y, \quad (1.5.14)$$

*where*

$$M = \sum_{i=1}^n \lambda(x_i) f(x_i) f'(x_i), \quad (1.5.15)$$

$$Y = \sum_{i=1}^n \lambda(x_i) y_i f(x_i);$$

(2) *an unbiased estimate for  $b^2$  will be*

$$\hat{b}^2 = (n - m)^{-1} \sum_{i=1}^n \lambda(x_i) [y_i - \hat{\theta}' f(x_i)]^2. \quad (1.5.16)$$

The proof is identical to the proof of the preceding theorem if we make the transformation

$$\tilde{f}(x) = \lambda^{1/2}(x) f(x), \quad \tilde{y}_i = \lambda^{1/2}(x_i) y_i, \quad (i = 1, 2, \dots, n).$$

When the function  $\lambda(x)$  is unknown, an estimate analogous to (1.5.3) and (1.5.14) cannot be constructed. In this case it is recommended to take several measurements at each point  $x_i$  in order to

permit the computation of an estimated value of the dispersion

$$b_t^2 = (n_t - 1)^{-1} \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2 \quad (1517)$$

If in formulas (138) and (139) or, whichever is more appropriate, in formulas (1337), we transform  $b_t^{-2}$  into  $\bar{b}_t^{-2}$ , then it is not difficult to show that the corresponding estimators will be close to the best linear estimators

### 1.6 Regression Analysis in the Presence of Errors in the Determination of the Control Variables

I. The mathematical apparatus for regression analysis above was developed under the assumption that the process of fixing the control variables  $x$  was deterministic. In practice, this requirement of determinism is very often not satisfied. If the uncertainty in fixing  $x$  is not large, then the usual methods of regression analysis can be used without leading to any essential error. In the contrary case, as will be shown in what follows, the values of the estimates of confidence intervals can be significantly different from the actual ones. In particular, the values of the parameters obtained by methods of least squares without weighting the errors in the determination of  $x$  can be either strongly biased or simply inconsistent.

Suppose we have the situation presented in the remarks following (121), with the exception of the fact that the control variables  $x$  are not deterministic, but random variables. We will assume, following [16], that the values of the coordinates of the control variables  $x$  are strictly constant in the course of the given measurements (for example, this can be the position of the measuring apparatus in space), but all that is known about these values is that they are chosen from some general collection with a density function  $p(x, x_0)$ , the first few moments of which are assumed known and the remaining are finite:

$$\begin{aligned} E(x_{\alpha}) &= x_{0\alpha}, \\ E[(x_{\alpha_1} - x_{0\alpha_1})(x_{\alpha_2} - x_{0\alpha_2})] &= d_{\alpha_1 \alpha_2}, \\ E[(x_{\alpha_1} - x_{0\alpha_1})(x_{\alpha_2} - x_{0\alpha_2}) \dots (x_{\alpha_K} - x_{0\alpha_K})] &= d_{\alpha_1 \alpha_2 \dots \alpha_K}, \\ d_i < \infty, \quad i > K \quad (\alpha_j = 1, 2, \dots, k) \end{aligned} \quad (161)$$

For the majority of practical problems it is completely sufficient to know only the values of the first and second moments.

Since the experimenter knows only the quantities  $y_i$  and  $x_{0i}$  ( $i = 1, 2, \dots, n$ ;  $n$  is the number of points at which measurements are taken), it is necessary to know the analytic form of the density function  $p(y, x_0)$  or at least its first and second moments for any value of  $x_0$  in order to construct a scheme of regression analysis. We will assume that the analytic form of the density function of the resulting observations  $p(y | x)$  for fixed  $x$  is known. From the formula of total probability (cf., e.g., [12]) it is not difficult to obtain that

$$p(y, x_0) = \int_{\Omega} p(y | x) p(x, x_0) dx, \quad (1.6.2)$$

where  $\Omega$  is the region of admissible values of  $x$  (for a given  $x_0$ ) and  $dx = dx_1 \cdot dx_2 \cdots dx_k$ . The integral (1.6.2), depending on the unknown parameters, permits analytical computation only for some trivial cases. Therefore it is significantly more useful to seek several of the first moments of  $p(y | x_0)$ . Since, in the basic scheme of regression analysis, the method of least squares will be used, we restrict ourselves to computation of only the first and second moments.

Corresponding to the definition and Eq. (1.6.2), the first moment equals

$$E(y, x_0) = \int_Y \int_{\Omega} y p(y | x) p(x, x_0) dx dy, \quad (1.6.3)$$

where  $Y$  is the region of permissible values of  $y$ . Interchanging the order of integration and making use of the fact that

$$\int_Y y p(y | x) dy = E(y | x) = \eta(x, \theta),$$

we obtain

$$E(y, x_0) = \int_{\Omega} \eta(x, \theta) p(x, x_0) dx. \quad (1.6.4)$$

If the function  $\eta(x, \theta)$  is smooth in a neighborhood of  $x_0$ , then it is possible to use its Taylor series expansion:

$$\eta(x, \theta) = \eta(x_0, \theta) + (\nabla_1 X_1) + \frac{1}{2}(\nabla_2 X_2) + \cdots, \quad (1.6.5)$$

where

$$(\nabla_j X_i) = \sum_{\alpha_1 \alpha_2 \cdots \alpha_j} \nabla_{j \alpha_1 \alpha_2 \cdots \alpha_j} X_{j \alpha_1 \alpha_2 \cdots \alpha_j},$$

and the elements of the matrices  $\nabla_j$  and  $X_j$  are defined as

$$\begin{aligned} X_{j_{0_1} \dots j_{0_k}} &= (x_{j_1} - x_{0j_1})(x_{j_2} - x_{0j_2}) \dots (x_{j_k} - x_{0j_k}) \\ \nabla_{j_{0_1} \dots j_{0_k}} &= \frac{\partial}{\partial x_{j_1}} \frac{\partial}{\partial x_{j_2}} \dots \frac{\partial}{\partial x_{j_k}} \eta(x, \theta) \Big|_{x=x_0} \end{aligned} \quad (166)$$

From (164), (165), and (161) it follows that

$$\begin{aligned} E(y, x_0) &= \eta(x_0, \theta) + \frac{1}{2}(\nabla_x^2 d_x) \\ &\quad + [\text{terms of the type } (1/k!)(\nabla_K d_K)] \end{aligned} \quad (167)$$

In the future, we will assume that terms with  $K > 2$  can be disregarded. In the majority of practical cases, for fairly small errors in the control variables and for sufficiently smooth functions  $\eta(x, \theta)$ , this assumption is satisfied. The computation of members of higher orders (under the assumption that the corresponding matrices of the  $K$ th moments of  $x$ ,  $K > 2$  are known) makes further computations only more cumbersome. Difficulties of a principal character do not arise. It must be pointed out that the computation of these members, as a rule, becomes necessary if  $x_0$  is found close to the boundary of its permissible values.

From expression (165) of the function  $\eta(x, \theta)$  it is not difficult to compute the second central moment

$$\begin{aligned} b^2(y, x_0) &= \int_Y [y - E(y, x_0)]^2 p(y, x_0) dy \\ &= b^2 + \int_D [\eta(x, \theta) - E(y, x_0)]^2 p(x, x_0) dx \end{aligned} \quad (168)$$

where  $b^2 = \int_Y [y - \eta(x, \theta)]^2 p(y | x) dy$  is the dispersion of the random variable  $y$  at the point  $x$ , in the absence of errors in the controlling variables. From (165), (167), and (168) it follows that

$$b^2(y, x_0) = b^2 + (\nabla_x^2 d_x) + O(\delta d_3) \quad (169)$$

where

$$\nabla_x^2 d_x = \frac{\partial \eta(x, \theta)}{\partial x_a} \Big|_{x=x_0} \quad \frac{\partial \eta(x, \theta)}{\partial x_b} \Big|_{x=x_0}$$

and

$$\delta_{\alpha\beta\gamma} = \frac{\partial \eta(x, \theta)}{\partial x_\alpha} \Big|_{x=x_0} \cdot \frac{\partial^2 \eta(x, \theta)}{\partial x_\beta \partial x_\gamma} \Big|_{x=x_0},$$

The results obtained can be strengthened if we make supplementary assumptions about the form of the functions  $p(y | x)$  and  $p(x, x_0)$ . For example, when the distribution function  $p(x, x_0)$  is symmetric with respect to  $x_0$  we may obtain that

$$E(y, x_0) = \eta(x_0, \theta) + \frac{1}{2}(\nabla_2 d_2) + O[\nabla_4(d_2 \times d_2)], \quad (1.6.10)$$

and

$$b^2(y, x_0) = b^2(x_0) + (\nabla_1^2 d_2) + O[(\nabla_2 d_2)^2]. \quad (1.6.11)$$

The necessity of using approximation (1.6.9) or (1.6.11) [but not (1.6.7) or (1.6.10)] disappears if at each point  $x_{0i}$  the number of measurements is sufficient for finding a satisfactory estimate of  $b^2(y, x_{0i})$ . We consider an estimate  $\hat{b}^2(y, x_{0i})$  satisfactory if its dispersion satisfies the inequality

$$d[\hat{b}^2(y, x_{0i})] \ll (\nabla_{1i}^2 d_{2i}). \quad (1.6.12)$$

The value  $\nabla_{1i}^2$  depends on the true value  $\theta_0$  and *a priori* is unknown. Therefore, in practical computations the variables in (1.6.12) must be replaced with rough estimates  $\hat{\theta}_0$ , which can be obtained by carrying out a corresponding regression analysis without taking into account errors in the controlling variables.

II. We consider an iterative method for finding values for the parameters  $\theta$ . One of the advantages of this method is the possibility of setting up a program for an ECM,<sup>1</sup> a basic part of which will be the program of the usual regression analysis (without taking into account the errors in the controlling variables).

The iterative process is constructed in the following way.

1. The sum is set up

$$M = \sum_{i=1}^n [y_i - \eta(x_i, \theta) - \frac{1}{2}(\nabla_{2i} d_{2i})]^2 \overset{0}{w}_i, \quad (1.6.13)$$

<sup>1</sup> ECM means electronic computing machine

where

$$\overset{0}{w}_i = b_i^{-2} = b^{-2}(x_{0i})$$

The values  $\overset{0}{\theta}$  are sought for which the minimum of  $M$  is attained (for  $\overset{0}{w}_i$  constant)

2 The quantity  $(\nabla_{1i}^2(\overset{0}{\theta}) d_{2i})$  is evaluated

3 The sum is computed

$$\overset{1}{M} = \sum_{i=1}^n [y_i - \eta(x_i, \overset{0}{\theta}) - \frac{1}{2}(\nabla_{2i} d_{2i})]^2 \overset{1}{w}_i, \quad (1.6.14)$$

where  $\overset{1}{w}_i = [b_i^{-2} + (\nabla_{1i}^2 d_{2i})]^{-1}$ . The value  $\overset{1}{\theta}$  is found for which the minimum of  $M_1$  is attained (for  $\overset{1}{w}_i$  constant). Operations 2 and 3 are repeated, respectively, for  $\overset{1}{\theta}$ ,  $\overset{2}{\theta}$ , and so on, as long as the process does not converge. The computation stops if, for example,

$$\max_{\alpha} |\overset{\alpha}{\theta}_s - \overset{\alpha}{\theta}_{s-1}| \leq \epsilon \quad (\alpha = 1, 2, \dots, m) \quad (1.6.15)$$

where  $\epsilon$  is a small positive number given beforehand. The value  $\overset{s}{\theta}$ , for which (1.6.15) is satisfied is taken as the value of the estimator  $\overset{s}{\theta}$  for the unknown parameter  $\theta$ .

For seeking the minimum of a quadratic form  $M$  it is possible to make use of the program of the usual method of least squares (MLS).

Sometimes it is convenient at each  $s+1$ st stage to consider as constants in the summation not only  $w_i$  but also the quantity  $\frac{1}{2}(\nabla_{2i} d_{2i})$ , which correspondingly assumes the values  $\frac{1}{2}[\nabla_{2i}(\overset{s}{\theta}) d_{2i}]$ . The iterative process in this case will converge somewhat more slowly but the region of its convergence remains the same.

**III** We investigate the statistical properties of the estimator  $\overset{s}{\theta}$ . We consider the case of linear parametrization  $\eta(x, \theta) = \theta f(x)$ . In this case the  $(s+1)$ st approximation can be represented in the form

$$\overset{s+1}{\theta} = [F^T w F]^{-1} F^T w y \quad (1.6.16)$$

Here

$$\begin{aligned} F &= \|\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n)\|, \\ \varphi_\alpha(x_i) &= (\partial/\partial\theta_\alpha)[\eta(x, \theta) + \frac{1}{2}(\nabla_{2i} d_{2i})] \quad (\alpha = 1, 2, \dots, m), \\ \overset{s}{w}_{ii} &= \overset{s}{w}_i = [b_i^2 + (\nabla_{1i}^2(\theta) d_{2i})]^{-1}, \quad w_{ij} = 0 \quad (i \neq j), \\ y' &= \|y_1, y_2, \dots, y_n\|. \end{aligned} \quad (1.6.17)$$

If the iterative process (1)–(3) converges, then its point of convergence  $\hat{\theta}$  is the solution to the equation [15]:

$$\theta = [Fw(\theta)F']^{-1}Fw(\theta)y. \quad (1.6.18)$$

It is not difficult to verify that (1.6.18) is equivalent to the equation

$$\sum_{i=1}^n [y_i - \varphi'(x_i)\theta] \varphi(x_i) w_i(\theta) = 0. \quad (1.6.19)$$

We note that the estimate  $\hat{\theta}$  does not minimize the quadratic form  $S = \sum_{i=1}^n w_i(\theta)[y_i - \theta'\varphi(x_i)]^2$ , that is, it is not a MLS estimate with the expansions (1.6.7) and (1.6.9). Indeed,

$$\begin{aligned} (\partial/\partial\theta_\alpha) \sum_{i=1}^n w_i(\theta)[y_i - \theta'\varphi(x_i)]^2 &= -2 \sum_{i=1}^n [y_i - \theta'\varphi(x_i)] \varphi_\alpha(x_i) w_i \\ &\quad + \sum_{i=1}^n [y_i - \theta'\varphi(x_i)]^2 (\partial w_i / \partial \theta_\alpha), \\ &\quad (\alpha = 1, 2, \dots, m). \end{aligned} \quad (1.6.20)$$

However, (1.6.19) only makes the first term in (1.6.20) zero. Moreover, it is possible to show that the estimate corresponding to  $\min_\theta S(\theta)$  in the general case is not consistent.

**Theorem 1.6.1.** *If*

$$\lim_{n \rightarrow \infty} n^{-1} Fw(\theta_0) F' = M(\theta_0) \quad \text{and} \quad |M(\theta_0)| \neq 0,$$

*then the estimate  $\hat{\theta}$ , which is the solution of Eq. (1.6.18) is consistent [in the framework of approximation (1.6.7)] and*

$$\lim_{n \rightarrow \infty} nD(\hat{\theta}) = M^{-1}(\theta_0). \quad (1.6.21)$$

*Proof* For the purpose of minimizing algebraic manipulations, we carry out the proof of the theorem for  $m = 1$ . Generalizing to the case of a larger number of parameters will be obvious.

We set up the quantity  $q_i(\theta) = [\theta\varphi(x_i) - y_i]\varphi(x_i)w_i(\theta)$  in a neighborhood of the true value  $\theta_0$  in the form

$$q_i(\theta) = q(\theta_0) + q_i(\theta_0)(\theta - \theta_0) + \frac{1}{2}q''(\theta_0)(\theta - \theta_0)^2, \quad (1.6.22)$$

where  $\theta_0'$  is located between  $\theta$  and  $\theta_0$ ,

$$q(\theta) = [\partial q(\theta)/\partial\theta] = \varphi^2(x_i)u(\theta) + [\theta\varphi(x_i) - y_i]\varphi(x_i)[\partial w_i(\theta)/\partial\theta],$$

$$q_{ii}(\theta) = [\partial^2 q_i(\theta)/\partial\theta^2] = 2\varphi^2(x_i)[\partial w_i(\theta)/\partial\theta] + [\theta\varphi(x_i) - y_i]\varphi(x_i)[\partial^2 w_i(\theta)/\partial\theta^2]$$

Multiplying both sides of (1.6.22) by  $n^{-1}$  and summing in  $i$ , for  $\theta = \hat{\theta}$  we obtain

$$\begin{aligned} n^{-1} \sum_{i=1}^n q_i(\hat{\theta}) &= n^{-1} \sum_{i=1}^n q_i(\theta_0) + n^{-1} \sum_{i=1}^n q_i(\theta_0)(\hat{\theta} - \theta_0) \\ &\quad + (2n)^{-1} \sum_{i=1}^n q''(\theta_0)(\hat{\theta} - \theta_0)^2 = 0 \end{aligned} \quad (1.6.23)$$

From the strong law of large numbers [10] it follows that there is an  $N$  such that for  $n \geq N$

$$P_1 \left( \left| n^{-1} \sum_{i=1}^n q_i(\theta_0) \right| \geq \delta \right) < \frac{1}{2}\epsilon,$$

$$P_2 \left( \left| n^{-1} \sum_{i=1}^n q_i(\theta_0) - n^{-1} \sum_{i=1}^n \varphi^2(x_i - \theta_0)w_i(\theta_0) \right| \geq \delta \right) < \frac{1}{2}\epsilon, \quad (1.6.24)$$

$$P_3 \left( \left| n^{-1} \sum_{i=1}^n q''(\theta_0) \right| \geq K \right) < \frac{1}{2}\epsilon, \quad K < \infty$$

Relying on (1.6.23) and (1.6.24) and repeating word-for-word the standard arguments of the first part of Theorem 33.3 from [II], we can easily show that the inequality

$$P(|\hat{\theta} - \theta_0| \leq \delta) > 1 - \epsilon \quad (1.6.25)$$

is satisfied for all  $n \geq N$ . In this way, the estimator is consistent

We compute the dispersion of  $\hat{\theta}$  for large  $n$ . Since the estimator is consistent, as  $n \rightarrow \infty$

$$\sum_{i=1}^n q_i(\hat{\theta}) = \sum_{i=1}^n q_i(\theta_0) + \sum_{i=1}^n \dot{q}_i(\theta_0)(\hat{\theta} - \theta_0) + O[(\hat{\theta} - \theta_0)^2] = 0, \quad (1.6.26)$$

or with accuracy up to terms of the second order,

$$\hat{\theta} - \theta_0 = - \sum_{i=1}^n q_i(\theta_0) / \sum_{i=1}^n \dot{q}_i(\theta_0). \quad (1.6.27)$$

Taking the dispersion of both sides of (1.6.27) and using the strong law of large numbers, we obtain

$$\lim_{n \rightarrow \infty} \left| D(\hat{\theta}) - \sum_{i=1}^n w_i \varphi^2(\theta_0, x_i) \right| = 0. \quad (1.6.28)$$

The theorem is proved.

From (1.6.21) and the consistency of the estimator  $\hat{\theta}$ , it follows that the dispersion matrix  $D(\hat{\theta})$  can be computed by the approximation formula

$$D(\hat{\theta}) \simeq \left[ \sum_{i=1}^n w_i(\hat{\theta}) \varphi(\hat{\theta}, x_i) \varphi'(\hat{\theta}, x_i) \right]^{-1}. \quad (1.6.29)$$

**IV.** We investigate the question of convergence of the iterative method of solving Eq. (1.6.18). For simplicity, as before, we will consider the case of one parameter. If  $\overset{0}{\theta}$  lies in a sufficiently small neighborhood of  $\hat{\theta}$ , then, as is known [15], for the convergence of the iterative method of solving Eq. (1.6.18) it is sufficient to have the inequality

$$\left| \frac{\partial L(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} < 1, \quad (1.6.30)$$

where

$$L(\theta) = [Fw(\theta) F']^{-1} Fw(\theta) y. \quad (1.6.31)$$

By an immediate differentiation it is easy to verify that

$$\partial L(\theta)/\partial \theta = [Fw(\theta) F']^{-1} F [\partial w(\theta)/\partial \theta] [y - F'\theta]. \quad (1.6.32)$$

From (1.6.32) it is clear that the derivative  $\partial L(\theta)/\partial \theta$  depends on the results of the observations  $y$  and therefore is a random variable.

Since the estimator  $\hat{\theta}$ , which is an exact solution of Eq. (1.6.18), is consistent, then for sufficiently large  $n$ , inequality (1.6.30) can be transformed to

$$\{ \partial L(\theta) / \partial \theta \}_{\theta=\theta_0} < 1 \quad (1.6.33)$$

From (1.6.7) and (1.6.32) it follows that

$$E[\partial L(\theta) / \partial \theta |_{\theta=\theta_0}] = 0 \quad (1.6.34)$$

In this way, for all possible choices  $y' = [y_1, y_2, \dots, y_n]$  the iterative process under consideration converges in the mean.

However, there exist samples for which the iterative process may not converge. From Chebyshev's inequality (cf., e.g., [10, 11]) and (1.6.32) it is not difficult to write an approximate upper bound for the probability  $P$  of obtaining such samples

$$P[|\partial L(\theta) / \partial \theta|_{\theta=\theta_0} \geq 1] \leq d(\theta_0), \quad (1.6.35)$$

where

$$d(\theta) = [Fw(\theta) F]^{-1} F[\partial w(\theta) / \partial \theta] w^{-1}(\theta) [\partial w(\theta) / \partial \theta] F^* [Fw(\theta) F]^{-1}$$

The upper bound given by (1.6.35) is very coarse, and knowing the distribution law of the observations makes it possible to lower this bound significantly. Thus, for example, if the resulting observations follow the normal distribution, this bound is equal to

$$P[|\partial L(\theta) / \partial \theta|_{\theta=\theta_0} \geq 1] = 1 - 2\varphi[d^{1/2}(\theta_0)] \quad (1.6.36)$$

where  $\varphi(x)$  is the normal distribution function with parameters  $(0, 1)$ .

Analogous results are easy to obtain for the case in which there are several unknown parameters.

In Parts III-IV of this section the calculations were carried out for  $\eta(x, \theta) = \theta f(x)$ . The results are easily generalized to the case where  $\eta(x, \theta)$  is of an arbitrary form. For this in all final formulas it is sufficient to set

$$f_a(x) = \partial \eta(x, \theta) / \partial \theta_a |_{\theta=\theta_0}, \quad (1.6.37)$$

and for constructing the iterative process, for each step  $s$

$$j_a(x) = \partial \eta(x, \theta) / \partial \theta_a |_{\theta=\theta_s} \quad (1.6.38)$$

V. We estimate the value of the error in  $x$  for which it is possible to apply the usual method of least squares (cf. Sections 1.2 and 1.4).

From (1.6.14) it follows that  $\overset{0}{\theta} \simeq \overset{1}{\theta}$  (i.e., application of the iterative method is not necessary), if

$$\max_i [(\nabla_{1i}^2 d_{2i})/b_i^2] \ll 1 \quad \text{and} \quad \max_i |(\nabla_{2i} d_{2i})/2\eta(x_i, \overset{0}{\theta})| \ll 1,$$

or more roughly

$$\sum_{\alpha=1}^k \left( \frac{\partial \eta(x, \overset{0}{\theta})}{\partial x_\alpha} \right)^2 \Big|_{x=x_i} d_{2\alpha\alpha} b_i^{-2} \ll 1$$

and

$$\sum_{\alpha=1}^k \frac{\partial^2 \eta(x, \overset{0}{\theta})}{\partial x_\alpha^2} \Big|_{x=x_i} d_{2\alpha\alpha} \eta^{-1}(x_i, \overset{0}{\theta}) \ll 1$$

for all points where the measurements are taken. In other words, the smaller the first and second derivatives, the smaller the error in  $x$  allowed in the analysis of experimental data by the usual methods of regression analysis.

### 1.7. Analysis of Experimental Data in the Case of Simultaneous Observations of Several Variables

Very often, for one of the values of the control variables  $x$  it is possible to measure several variables  $y_1, y_2, \dots, y_l$ . Chemical experiments can serve as the simplest example of this in which several quantities are formed for which the concentration of each of them can be measured.

The variables  $y_1, y_2, \dots, y_l$  can be correlated. We will assume that their dispersion matrix is known:

$$D(y|x) = \begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ b_{l1} & b_{l2} & \cdots & b_{ll} \end{vmatrix}, \quad (1.7.1)$$

where  $y' = \|y_1, y_2, \dots, y_l\|$ .

Suppose it is also known that

$$E(y | x) = \begin{bmatrix} \eta_1(x, \theta) \\ \eta_2(x, \theta) \\ \vdots \\ \eta_l(x, \theta) \end{bmatrix} \quad (1.7.2)$$

We consider the case of linear parametrization, i.e.,

$$\eta_k(x, \theta) = \theta' f_k(x) \quad (k = 1, 2, \dots, l) \quad (1.7.3)$$

We introduce a notation

$$W_i = D^{-1}(y | x_i), \quad F(x_i) = [f_1(x_i) \ f_2(x_i) \ \dots \ f_l(x_i)] \quad (1.7.4)$$

The following theorem is valid

**Theorem 1.7.1.** *The best linear estimator for  $\theta$  is*

$$\hat{\theta} = M^{-1}Y,$$

where

$$M = \sum_{i=1}^n F(x_i) W_i F'(x_i) \quad \text{and} \quad Y = \sum_{i=1}^n F(x_i) W_i y_i \quad (1.7.5)$$

The dispersion matrix of the estimator  $\hat{\theta}$  is equal to

$$D(\hat{\theta}) = M^{-1} \quad (1.7.6)$$

The theorem is proved analogously to Theorem 1.3.2. Changing in the proof  $f(x_i)$  to  $F(x_i)$  and  $b_i^{-2}$  (or  $w_i$ ) to  $D(y | x_i)$  (or  $W_i$ ) and repeating its proof word-for-word we obtain the resulting Theorem 1.7.1

It is not difficult to show that all the corollaries of Theorem 1.3.2 hold in this case. For our purposes in what follows, two corollaries, which are obvious generalizations of Corollaries 4 and 5 of Theorem 1.3.2, will be necessary

**Corollary 1.** *The best linear estimator of the function  $\eta_k(x, \theta)$  ( $k = 1, 2, \dots, l$ ) is the function*

$$\hat{\eta}_k(\tau) = \hat{\theta} f_k(\tau) \quad (1.7.7)$$

The dispersion matrix of the estimator  $\hat{\eta}_h(x)$  is equal to

$$d(x) = F'(x) M^{-1} F(x). \quad (1.7.8)$$

**Corollary 2.** If at the point  $x_i$  ( $i = 1, 2, \dots, n$ ) the results  $y_{i1}, y_{i2}, \dots, y_{in_i}$  were obtained, then the best linear estimate for  $\theta$  would be

$$\hat{\theta} = M^{-1} Y, \quad (1.7.9)$$

where

$$M = \sum_{i=1}^n F(x_i) W_i F'(x_i), \quad (1.7.10)$$

$$Y = \sum_{i=1}^n F(x_i) W_i y_i,$$

and

$$W_i = n_i D^{-1}(y_i),$$

$$y_i = n_i^{-1} \sum_{j=1}^n y_{ij}. \quad (1.7.11)$$

Here we assume that  $D(y_j | x_i)$  is the same for all  $j = 1, 2, \dots, n_i$ .

We will not write out the remaining formulas which are generalizations to the results of Sections 2–5. The easy computations necessary for obtaining them are left to the reader.

## 1.8. Methods of Comparing Results of Experiments

In this and the following sections, all arguments are carried out under the assumption that at a given point of the factor space only one quantity can be measured. Generalization to the case when it is possible to have simultaneous measurements of several quantities  $y_1, y_2, \dots, y_l$  is obvious and is partially carried out in Chapter 5.

First of all, we make precise the concept of an experiment. We will call the collection of variables

$$\begin{array}{lll} y_{11}', y_{12}', \dots, y_{1n_1}'; & y_{21}', y_{22}', \dots, y_{2n_2}'; & \dots; & y_{n1}', y_{n2}', \dots, y_{nn_n}'; \\ w_1' & ; & w_2' & ; \dots; & w_n' \\ x_1' & ; & x_2' & ; \dots; & x_n' \end{array} \quad (1.8.1)$$

an experiment  $\mathcal{E}_j$ . We recall that  $w_i' = n_i(b_i')^{-2}$ .

We will say that the experiment  $\mathcal{E}_1$  is distinct from the experiment  $\mathcal{E}_2$  (abbreviated  $\mathcal{E}_1 \neq \mathcal{E}_2$ ) if at least one of the variables for  $j = 1$  is distinct from the corresponding variable for  $j = 2$ . Generally speaking for distinct experiments  $\mathcal{E}_j$ , the values of the estimator  $\hat{\theta}_j$  [or  $\hat{\eta}_j(x)$ ] will be distinct as variables and also as "approximations" of them to their true values. Before we go on to a design, that is, to a choice of the best experiment, it is necessary to have criteria for comparing experiments. It is natural to consider experiment  $\mathcal{E}_1$  as preferable to experiment  $\mathcal{E}_2$  ( $\mathcal{E}_1 > \mathcal{E}_2$ ) if the value of the estimate  $\hat{\theta}(\mathcal{E}_1)$  [or  $\hat{\eta}(x, \mathcal{E}_1)$ ] is "closer" to the true value  $\theta_0$  [or  $\eta_0(x)$ ], than the value  $\hat{\theta}(\mathcal{E}_2)$  [or  $\hat{\eta}(x, \mathcal{E}_2)$ ].

The choice of the best linear estimate  $\hat{\theta}$  among the remaining estimators  $\hat{\theta}$  is based on a comparison of the dispersion matrices  $D(\hat{\theta})$  and  $D(\hat{\theta})$  or some combination of their elements. Therefore the estimators  $\hat{\theta}(\mathcal{E}_1)$  and  $\hat{\theta}(\mathcal{E}_2)$  can be conveniently compared by relying respectively, on the corresponding dispersion matrices  $D(\mathcal{E}_1)$  and  $D(\mathcal{E}_2)$  of the estimators of the parameter  $\theta$ .

We enumerate some of the most frequently required properties of a comparison of experiments.

(a) Experiment  $\mathcal{E}_1$  is preferred to experiment  $\mathcal{E}_2$  if the difference  $D(\mathcal{E}_2) - D(\mathcal{E}_1)$  is a positive-definite matrix, or  $\mathcal{E}_1 > \mathcal{E}_2$ , if

$$D(\mathcal{E}_1) < D(\mathcal{E}_2) \quad (1.8.2)$$

(b)  $\mathcal{E}_1 > \mathcal{E}_2$  if

$$|D(\mathcal{E}_1)| < |D(\mathcal{E}_2)| \quad (1.8.3)$$

(c)  $\mathcal{E}_1 > \mathcal{E}_2$  if

$$\text{Tr } D(\mathcal{E}_1) < \text{Tr } D(\mathcal{E}_2) \quad (1.8.4)$$

(d)  $\mathcal{E}_1 > \mathcal{E}_2$  if

$$\max_{\mathbf{x}} D_{\mathbf{x}\mathbf{x}}(\mathcal{E}_1) < \max_{\mathbf{x}} D_{\mathbf{x}\mathbf{x}}(\mathcal{E}_2) \quad (1.8.5)$$

(e)  $\mathcal{E}_1 > \mathcal{E}_2$  if

$$D(\varphi \cdot \mathcal{E}_1) < D(\varphi \cdot \mathcal{E}_2) \quad (1.8.6)$$

where

$$\varphi = L\theta \quad (1.8.7)$$

and

$$D(\varphi \cdot \mathcal{E}) = \Sigma D(\mathcal{E}) \Sigma \quad (1.8.8)$$

Comparison method (e) is a slight generalization of method (a) [(e) transforms into (a) when  $L = I_m$ ]. In an analogous manner it is possible to generalize comparison methods (b)–(d). For example,  $\mathcal{E}_1 > \mathcal{E}_2$  if

$$\text{Tr } D(\varphi, \mathcal{E}_1) < \text{Tr } D(\varphi, \mathcal{E}_2). \quad (1.8.9)$$

Sometimes only part of the parameters  $\theta_1, \theta_2, \dots, \theta_l$  ( $l < m$ ) of the entire collection  $m$  of unknown parameters are of particular interest to the experimenter. In this case, experiments are compared by the same methods used for  $l = m$  but instead of the full matrix  $D(\mathcal{E})$  it is necessary to use the submatrix

$$D_u(\mathcal{E}) = \begin{vmatrix} D_{11} & D_{12} & \cdots & D_{1l} \\ D_{21} & D_{22} & \cdots & D_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ D_{l1} & D_{l2} & \cdots & D_{ll} \end{vmatrix}.$$

In this case, if we are using (1.8.7) it is possible to set

$$L = \begin{vmatrix} \underbrace{1 & 0 & \cdots & 0}_{l}, & \underbrace{0 & 0 & 0}_{m-l} \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & & 0 & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & & 0 & 0 & 0 \end{vmatrix}.$$

The preference criteria determined by (1.8.2)–(1.8.9) compare experiments in the parameter space.

The criteria just enumerated are particularly obvious in light of the geometric interpretation of the dispersion matrix of the estimator of the unknown parameters. It is well known (cf., for example [11]) that for normally distributed observations the surfaces of constant value of the density function  $p(\theta | \theta_0)$  are ellipsoids, the characteristics of which are completely determined by elements of the matrix  $D(\mathcal{E})$ . In Fig. 1 the ellipse for the two-dimensional case ( $m = 2$ ) is presented, in which  $p(\theta | \theta_0)$  is constant and with probability  $\sim 0.39$  covers the unknown value of the parameter  $\theta_0$  (the so-called ellipsoid of concentration). From the figure, the geometric interpretation of each of the given criteria for comparison of experiments is evident.

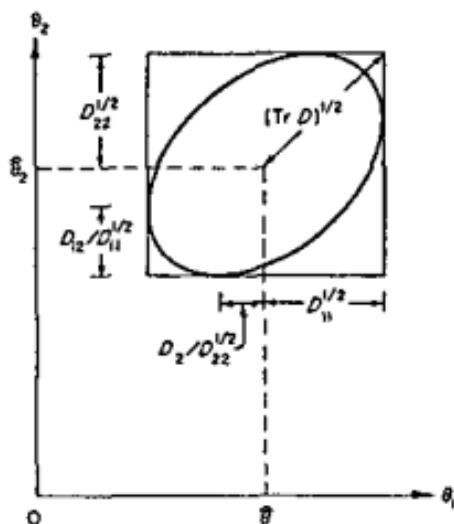


Fig. 1. Ellipsoid of concentration

We consider methods for comparison of experiments in the space of control variables

(f)  $\mathcal{E}_1 > \mathcal{E}_2$ , if for the dispersion of the respective estimators  $\hat{\eta}(x, \mathcal{E}_1)$  and  $\hat{\eta}(x, \mathcal{E}_2)$  the following inequality is satisfied

$$\max_{x \in Z} d(x, \mathcal{E}_1) < \max_{x \in Z} d(x, \mathcal{E}_2), \quad (1.8.10)$$

(g)  $\mathcal{E}_1 > \mathcal{E}_2$ , if

$$\int_Z d(x, \mathcal{E}_1) dx < \int_Z d(x, \mathcal{E}_2) dx \quad (1.8.11)$$

Usually the following cases are distinguished

(1) The region  $Z$  coincides with the region  $X$  of possible measurements

(2) The region  $Z$  is a subset of  $X$

(3) The region  $Z$  has no point in common with  $X$  (the problem of extrapolation)

The region  $Z$ , in particular, can be a single point. In this case, (1.8.10) becomes

$$d(x, \mathcal{E}_1) < d(x, \mathcal{E}_2), \quad (1.8.12)$$

where  $x$  can belong to  $X$ , or be located outside of it

EXAMPLE. We show that in the general case, an experiment  $\mathcal{E}_1$  can be preferred to an experiment  $\mathcal{E}_2$  in the sense of one or several criteria of comparison, and at the same time, under other criteria, experiment  $\mathcal{E}_2$  is preferred.

Let

$$E(y | x) = \theta_1 + \theta_2 x + \theta_3 x^2$$

and let there be two experiments with the same number of observations

$$\mathcal{E}_1 = \left\{ \begin{array}{l} y_1^1; \\ w_1^1 = 4; \\ x_1^1 = -1; \end{array} \quad \begin{array}{l} y_2^1; \\ w_2^1 = 4; \\ x_2^1 = 0; \end{array} \quad \begin{array}{l} y_3^1 \\ w_3^1 = 4 \\ x_3^1 = 1 \end{array} \right\},$$

$$\mathcal{E}_2 = \left\{ \begin{array}{l} y_1^2; \\ w_1^2 = 3; \\ x_1^2 = -1; \end{array} \quad \begin{array}{l} y_2^2; \\ w_2^2 = 6; \\ x_2^2 = 0; \end{array} \quad \begin{array}{l} y_3^2 \\ w_3^2 = 3 \\ x_3^2 = 1 \end{array} \right\}.$$

Here  $y_i^j$  is the mean value of the observations at the point  $x_i$  in experiment  $\mathcal{E}_j$ , and  $b^2 = 1$ .

Using (1.3.9), it is not difficult to obtain that for the first experiment

$$D(\mathcal{E}_1) = \frac{1}{12} \begin{vmatrix} 3 & 0 & -3 \\ 0 & \frac{3}{2} & 0 \\ -3 & 0 & \frac{9}{2} \end{vmatrix},$$

and for the second

$$D(\mathcal{E}_2) = \frac{1}{12} \begin{vmatrix} 2 & 0 & -2 \\ 0 & 2 & 0 \\ -2 & 0 & 4 \end{vmatrix}.$$

From this, after a simple computation we obtain

$$|D(\mathcal{E}_1)| = 6.75 (12^{-3}) \quad \text{and} \quad |D(\mathcal{E}_2)| = 8 (12^{-3}),$$

that is  $\mathcal{E}_1 > \mathcal{E}_2$  in the sense of criterion (b).

On the other hand,  $\text{Tr } D(\mathcal{E}_1) = \frac{3}{4}$  and  $\text{Tr } D(\mathcal{E}_2) = \frac{2}{3}$ , and the experiment  $\mathcal{E}_2$  in this manner is preferred to the experiment  $\mathcal{E}_1$  in the sense of criterion (c).

We turn our attention now to consideration of methods for comparing quantities not depending on the observations

### 1.9. The Loss Function for Regression Experiments

The experiments discussed in Section 1.8 were compared only by the final results. No attention was paid to the methods, the time, and the monetary or material losses that came about. In experimental practice one is far from indifferent to the losses obtained by this or that method. By loss  $\tau$ , here and in what follows we will understand the loss in experimental time, money, or material means necessary for conducting the experiment, and a series of other factors which can be characterized as the general cost of an experiment expressed, for example, in monetary units. Sometimes the experiment  $\mathcal{E}_1(\tau)$  is somewhat worse in the sense of (1.8.2)–(1.8.12) than the experiment  $\mathcal{E}_2(\tau)$ , but it can be significantly better if  $\tau_1 \ll \tau_2$  (the loss in the first experiment is much smaller than the loss in the second experiment). Therefore the choice of one or another experiment must be based on a comparison of some quantity  $\mathcal{R}(\mathcal{E})$  which depends on the loss  $\tau$ , and also on the value of the quantities (1.8.2)–(1.8.12). In the general case, the function  $\mathcal{R}(\mathcal{E})$  can be represented in the form

$$\mathcal{R}(\mathcal{E}) = \tau + \Psi[D(\mathcal{E})] \quad (1.9.1)$$

where  $\tau$  is the loss for experiment  $\mathcal{E}$ ,  $\Psi$  is some functional depending on the dispersion matrix of the estimates of the parameters. The function  $\mathcal{R}(\mathcal{E})$  is called the loss function for the experiment  $\mathcal{E}$ . The best experiment will be an experiment  $\mathcal{E}$  minimizing  $\mathcal{R}(\mathcal{E})$ . The functional  $\Psi$  is chosen depending only on the elements of the matrix  $D(\mathcal{E})$  because we rely only on the matrix  $D(\mathcal{E})$  for a description of the accuracy of the results of an experiment. The value  $\Psi$  characterizes the possible losses conditioned on an imprecise determination of  $\theta$  [or  $\hat{\eta}(x)$ ]. An analytic form of the functional is determined by the end-goal of the experiment. Usually  $\Psi$  is considered equal to

$$\Psi[D(\mathcal{E})] = k \mathcal{L}[D(\mathcal{E})] \quad (1.9.2)$$

where  $k$  is a constant normalizing multiplier and  $\mathcal{L}[D(\mathcal{E})]$  is determined by the method of comparison of experiments which is used in the given experimental situation. If the experiments are compared

by one of the methods (a)–(f) then the functional  $\mathcal{L}(\mathcal{E})$  can have the form

$$\mathcal{L}[D(\mathcal{E})] = |D(\mathcal{E})|, \quad (1.9.3)$$

$$\mathcal{L}[D(\mathcal{E})] = \text{Tr } D(\mathcal{E}), \quad (1.9.4)$$

$$\mathcal{L}[D(\mathcal{E})] = \max_{\alpha} D_{\alpha\alpha}(\mathcal{E}) \quad (\alpha = 1, 2, \dots, m), \quad (1.9.5)$$

$$\mathcal{L}[D(\mathcal{E})] = L'D(\mathcal{E})L, \quad (1.9.6)$$

where  $L' = \|l_1, l_2, \dots, l_m\|$ ;

$$\mathcal{L}[D(\mathcal{E})] = \max_x d(x, \mathcal{E}), \quad (1.9.7)$$

$$\mathcal{L}[D(\mathcal{E})] = \int_Z d(x, \mathcal{E}) dx; \quad (1.9.8)$$

the region  $Z$  is defined as in the explanation to (1.8.10) and (1.8.11).

If the experimenter is interested in  $l$  of the parameters from the  $m$ , then in (1.9.3)–(1.9.5) the matrix  $D(\mathcal{E})$  must be changed to the corresponding submatrix  $D_{ll}(\mathcal{E})$ .

The functional form (1.9.3)–(1.9.8) is convenient from the point of view of constructing future mathematical apparatus, and at the same time is suitable for the majority of practical problems. In the text we shall meet various other functionals  $\mathcal{L}$ , which are close in nature to the one just introduced.

In the future, the loss  $\tau$  will be assumed to be proportional to the cost of each distinct measurement:

$$\tau = \sum_{i=1}^n c_i n_i, \quad (1.9.9)$$

where  $c_i$  is the cost of an observation taken at the point  $x$ . The introduction of supplementary factors on which the function  $\tau$  can depend complicates its analytic form, makes the apparatus of planning experiments significantly more cumbersome, and only in the rarest cases gives a notably better description of reality. Typical dependence of the loss function  $\mathcal{R}$  on the number of observations is presented in Fig. 2.

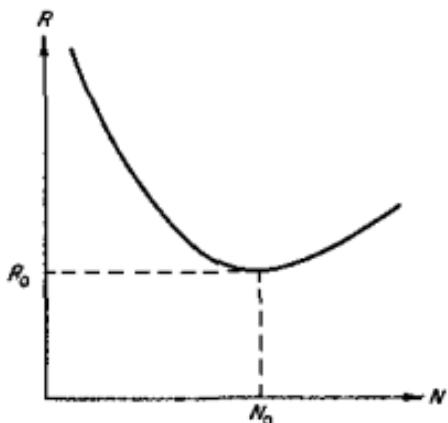


Fig. 2. Dependence of the loss function on the number of observations

### 1.10. The Concept of Experimental Design. Continuous Normalized Designs

I. In the general case [the function  $\eta(x, \theta)$  is arbitrary], the quantity  $\mathcal{R}(\xi)$  depends on all of the characteristics of the experiment  $\mathcal{E}$ . However, under linear parametrization [ $\eta(x, \theta) = \theta f(x)$ ] the loss function depends only on the quantities  $n_i, x_i$  ( $i = 1, 2, \dots, n$ ) and does not depend on the results of the measurements  $y$ , nor on the unknown values of the sought parameters  $\theta$ .

Indeed, corresponding to Theorem 1.3.2 and its Corollary 6, the dispersion matrix

$$D(\mathcal{E}) = M^{-1},$$

where

$$M = \sum_{i=1}^n w_i f(x_i) f'(x_i)$$

Since all of the expressions (1.9.3)–(1.9.8) depend only on the elements of the matrix  $D(\mathcal{E})$  and the function  $f(x)$ , the loss function depends only on the values  $n_i, x_i$  ( $i = 1, 2, \dots, n$ ).

The collection of variables

$$x_1, x_2, \dots, x_n; \quad n_1, n_2, \dots, n_n, \quad (1.10.1)$$

$$\sum_{i=1}^n n_i = N$$

is called the design of an experiment  $\mathcal{E}(N)$ . The points  $x_1, x_2, \dots, x_n$  are called the spectrum of the design  $\mathcal{E}(N)$ .

The concept of a normalized design will be useful to us. A normalized design  $\epsilon(N)$  is the collection of variables

$$p_1, p_2, \dots, p_n, \quad x_1, x_2, \dots, x_n, \quad (1.10.2)$$

where

$$\sum_{i=1}^n p_i = 1 \quad \text{and} \quad p_i = n_i/N.$$

The elements of the information matrix  $M[\mathcal{E}(N)]$  can be directly expressed by the quantities  $p_i$  ( $i = 1, 2, \dots, n$ ) and  $N$ :

$$M[\mathcal{E}(N)] = N \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i)$$

or

$$M[\mathcal{E}(N)] = NM[\epsilon(N)], \quad (1.10.3)$$

where for linear dependence of the response surface on the unknown parameters

$$M[\epsilon(N)] = \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i), \quad (1.10.4)$$

$$D[\mathcal{E}(N)] = N^{-1} D[\epsilon(N)]. \quad (1.10.5)$$

For a given normalized design  $\epsilon(N)$  the loss function has the form

$$\mathcal{R} = N \sum_{i=1}^n c_i p_i + kN^{-m} |D[\epsilon(N)]| \quad (1.10.6)$$

for the case (1.9.3); in the remaining case

$$\mathcal{R} = N \sum_{i=1}^n c_i p_i + kN^{-1} \mathcal{L}\{D[\epsilon(N)]\}, \quad (1.10.7)$$

where  $\mathcal{L}\{D[\epsilon(N)]\}$  is defined by one of the formulas (1.9.3)–(1.9.8) with  $D(\mathcal{E})$  replaced in them, respectively, by  $D[\epsilon(N)]$ .

Generalization of (1.10.6) and (1.10.7) to the case when the interest lies only in part of the parameters ( $l < m$ ) is clear.

Having set up the loss function, we reduce the design of an experiment to some extremal problem, in particular, to seeking the minimum of the loss function in  $N$  and  $\epsilon(N)$

$$\min_{N, \epsilon(N)} \mathcal{R}[N, \epsilon(N)] \quad (1.10.8)$$

The solution of the given problem is possible in principle, but the amount of computation rapidly grows as the number of unknown parameters and dimension of the space of control variables increases.

Under some supplementary assumptions the solution of problem (1.10.8) exists and very often simplifies, as we shall see later, to the use of some suitable tables.

II. Let  $N$  be given, i.e., the experimenter cannot obtain a conclusion from a smaller number of observations and at the same time cannot exceed a given level of loss.

We consider the case when the cost of observations does not depend on  $x_i$  ( $c_i = \text{const}$ ). In this case the loss function will have the form

$$\mathcal{R}[\mathcal{E}(N)] = Nc + k\mathcal{L}\{D[\mathcal{E}(N)]\} \quad (1.10.9)$$

and the design of the experiment will consist of seeking [compare with (1.10.8)]

$$\min_{\mathcal{E}(N)} \mathcal{R}[\mathcal{E}(N)] \quad (1.10.10)$$

From (1.10.9) it is not difficult to see that the search for the minimum (1.10.10) is equivalent to the search for

$$\min_{\mathcal{E}(N)} \mathcal{L}\{D[\mathcal{E}(N)]\} \quad (1.10.11)$$

**EXAMPLE** We consider the linear regression on the interval

$$E[\beta(x)] = \theta_1 + \theta_2x$$

in which  $D[\beta(x)] = 1$  and the measurements are possible in  $-1 \leq x \leq 1$ .

Let the number of observations be given and  $N = 3$ . It is evident that the observations can be distributed on no more than three points.

Corresponding to (1.3.7)

$$M[\mathcal{E}(3)] = \begin{vmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{vmatrix}$$

and it follows that

$$|M[\mathcal{E}(3)]| = 3(x_1^2 + x_2^2 + x_3^2) - (x_1 + x_2 + x_3)^2.$$

It is not difficult to verify that the maximum  $|M[\mathcal{E}(3)]|$  (minimum  $|D[\mathcal{E}(3)]|$ ) is attained for

$$x_1 = x_2 = -1 \quad \text{and} \quad x_3 = 1 \quad \text{or} \quad x_1 = x_2 = 1 \quad \text{and} \quad x_3 = -1.$$

In this case  $|M[\mathcal{E}(3)]| = 8$ . It is easy to obtain that for a given spectrum the design  $\mathcal{E}(3)$  also minimizes  $\text{Tr } D[\mathcal{E}(3)]$ .

The example considered is one of the simplest (two parameters and one factor). In the more complicated case, the direct search for the minimum  $\mathcal{L}\{D[\mathcal{E}(N)]\}$  is a much more difficult undertaking, since the dimension of the space in which it is necessary to conduct this search grows very rapidly (cf. Chapter 4)

**III.** In many experimental investigations, an “accuracy” of determination of the estimates of the sought parameters is given beforehand. In such situations it is necessary to deal, for example, with the definition of fundamental constants in physics and chemistry, in the construction of various technical tables, etc.

The given “accuracy” corresponds from the mathematical viewpoint to the given quantity  $\mathcal{L}\{D[\mathcal{E}(N)]\} = \mathcal{L}_0$ . The design of an experiment in this case consists of finding the minimum of the loss function in the design  $\mathcal{E}(N)$  and the number of observations  $N$  for the given  $\mathcal{L}_0$ .

If the cost of observations is the same for all points  $x_i$ , then the given problem is reduced to a sequential search for the minimum (1.10.11) which continues until one cannot find a design  $\mathcal{E}(N+1)$  for which

$$\min_{\mathcal{E}(N)} \mathcal{L}\{D[\mathcal{E}(N)]\} > \mathcal{L}_0 \geq \min_{\mathcal{E}(N+1)} \mathcal{L}\{D[\mathcal{E}(N+1)]\}.$$

The experiment is then conducted according to the design  $\mathcal{E}(N+1)$ .

**III.** The strongest results can be obtained when  $N$  is so large that the loss function can be considered as continuous in  $N$ .

We introduce the concept of a continuous normalized design.

A continuous normalized design  $\epsilon$  is the collection of quantities

$$x_1, x_2, \dots, x_n \quad p_1, p_2, \dots, p_n \quad (1.10.12)$$

$$\sum_{i=1}^n p_i = 1,$$

where the variables  $p_i$  ( $i = 1, 2, \dots, n$ ) can take on any value included between 0 and 1. In the more general case, the collection of points  $x$  can coincide with the collection of all points belonging to some closed region  $\lambda$ . In this case a continuous normalized design will be characterized by some measure  $\xi(x)$  given on the region  $X$  and satisfying the conditions

$$\int_X d\xi(x) = 1 \quad \xi(x) \geq 0 \quad x \in \lambda \quad (1.10.13)$$

Generalizing (1.10.4) to the case (1.10.13), it is possible to write

$$M(\epsilon) = \int_X \lambda(x) f(x) f'(x) d\xi(x) \quad (1.10.14)$$

In the case of a continuous measure, (1.10.14) takes on the form

$$M(\epsilon) = \int_X \lambda(x) f(x) f'(x) p(x) dx \quad \int_X p(x) dx = 1 \quad (1.10.15)$$

For the purely discrete case (the measure is concentrated on a finite number of points) the function  $p(x)$  sometimes can be conveniently considered equal to

$$p(x) = \sum p_i \delta(x - x_i)$$

where  $\delta(x - x_i)$  is the Dirac function. The definition of the Dirac  $\delta$  function can be found for example, in [5].

The given measure  $\xi(x)$  or the density function  $p(x)$  describes the design  $\epsilon$ . It is clear that for  $N \gg n$

$$\min_{N \in \epsilon(N)} \mathcal{R}[N | \epsilon(N)] \approx \min_{N \in \epsilon} \mathcal{R}(N, \epsilon) \quad (1.10.16)$$

The question of accuracy of approximation of the equality will be considered in Chapter 4.

If the cost of all observations is the same, then it is not difficult to find the minimum of the right-hand side of (1.10.16).

Indeed, for a given design  $\epsilon$ , a loss function of the form (1.10.9), and  $\mathcal{L}(\epsilon) = |D(\epsilon)|$ ,

$$\partial \mathcal{R}(N, \epsilon) / \partial N = c - mkN^{-(m+1)} |D(\epsilon)| = 0. \quad (1.10.17)$$

The optimal number of observations  $N$  will correspond to the root (more precisely, to the one which is closest to an integer) of the equation

$$N = \left[ \frac{mk |D(\epsilon)|}{c} \right]^{1/(m+1)}. \quad (1.10.18)$$

For  $\mathcal{L}(\epsilon)$  equal to one of the quantities (1.9.4)–(1.9.8):

$$N = \left[ \frac{k\mathcal{L}[D(\epsilon)]}{c} \right]^{1/2}. \quad (1.10.19)$$

Formulas (1.10.18) and (1.10.19) are valid when, respectively,

$$mk |D(\epsilon)| \gg c, \quad k\mathcal{L}[M(\epsilon)] \gg c,$$

that is, the cost of each observation must be small.

The loss function for any  $N$  is a strictly increasing function of  $|D(\epsilon)|$  or  $\mathcal{L}[D(\epsilon)]$ . It follows that the minimum value of  $\mathcal{R}$  will be attained for some continuous normalized design which minimizes  $|D(\epsilon)|$  or, correspondingly,  $\mathcal{L}[D(\epsilon)]$ . The value  $\min_{\epsilon} |D(\epsilon)|$  or  $\min_{\epsilon} \mathcal{L}[D(\epsilon)]$  does not depend on  $N$ . This simple but very important fact will be considered in the future directly from the definition of continuous normalized design.

In this manner, the design of an experiment in the case under consideration reduces to the search for a design  $\epsilon$  minimizing  $|D(\epsilon)|$  or one of the quantities  $\mathcal{L}[D(\epsilon)]$ , followed by a computation of the optimal number of observations  $N$ .

For the majority of designs, minimization of this or any other normalized quantity (1.9.3)–(1.9.8) is designated in the mathematical literature by a special name.

For example, the design minimizing the determinant  $|D(\epsilon)|$  is called the  $D$ -optimal design; the design minimizing  $\max_x d(x, \epsilon) = \max_x f'(x) D(\epsilon) f(x)$  is called the minimax design; minimization of the mean of the normalized dispersion  $m^{-1} \text{Tr } D(\epsilon)$  is called the  $A$ -optimal design; minimization of  $\max_x D_{xx}(\epsilon)$  is called the minimax design in the space of parameters.

# 2

## Continuous Optimal Designs (Statistical Methods)

### 2.1. Basic Properties of the Information Matrix

I. In this and the next two chapters, the design of experiments will be considered under the assumption that

$$E(y | x) = \eta(x | \theta), \quad (2.1.1)$$

where the analytic form of the function  $\eta(x | \theta)$  is given.

In Chapters 2 and 3, we assume that the efficiency function of the experiment is known with accuracy up to a constant multiplier  $b^{-2}$  [cf. (1.5.12)]. As will become clear from what follows, all results obtained in these chapters do not depend on the value of this multiplier. Therefore, without loss of generality, it is possible to set

$$b^{-2} = 1 \quad \text{and} \quad u(x) = \lambda(x),$$

where  $\lambda(x)$  is a known function.

Generalization of the results of Chapters 2 and 3, and methods of sequential design (Chapter 4), to the case where at each point of the factor space, simultaneous observations of several quantities  $y = [y_1, y_2, \dots, y_l]$  are possible, will be discussed in Chapter 5.

II. We study the basic properties of the information matrix.

In this chapter, if it is not specifically stated, we shall consider only continuous normalized designs and assume that  $\eta(x, \theta) = \theta'f(x)$ . Generalizations of the results to the case of nonlinear parametrizations will be considered in Section 2.8 and the following chapters.

First, we shall introduce some definitions [18], which will be necessary for us to formulate and prove Theorem 2.1.2 on the basic properties of the information matrix.

We will indicate by  $S_n$  the Euclidean  $n$ -dimensional space of vectors  $s' = \|s_1, s_2, \dots, s_n\|$ , where each  $s_i$  ( $i = 1, 2, \dots, n$ ) is a real number.

**DEFINITION 1.** The collection of all  $s$ , for which  $(s - s_0)'(s - s_0) < \delta^2$ ,  $\delta > 0$ , is called the sphere of radius  $\delta$  with center at the point  $s_0$ .

**DEFINITION 2.** The point  $s \in S$  is called an interior point of the set  $S$  if there exists a sphere with center at the point  $s$  which is a subset of the set  $S$ .

**DEFINITION 3.** The point  $s$  is called a boundary point of the set  $S$  if any sphere with center at the point  $s$  contains points belonging and not belonging to the set  $S$ .

**DEFINITION 4.** The set  $S$ , all of whose points are interior, is called open. The set  $S$  is called closed if its complement  $\bar{S}$ , the set of all points  $s$  not belonging to  $S$ , is open.

It is clear that a set containing its entire boundary is closed.

**DEFINITION 5.** The set  $S$  is called convex if any point

$$s = \alpha s_1 + (1 - \alpha)s_2, \quad \text{where } s_1 \in S, s_2 \in S, \text{ and } 0 \leq \alpha \leq 1,$$

belongs to this set.

Convex sets, for example, are points, straight lines, spheres, hyperplanes, etc.

The set  $S^*$  of points

$$s^* = \sum_{i=1}^k \alpha_i s_i,$$

where

$$\sum_{i=1}^k \alpha_i = 1, \quad \alpha_i \geq 0, \quad s_i \in S \quad (i = 1, 2, \dots, k; k = 1, 2, \dots),$$

is a convex set, as it is not difficult to show.

**DEFINITION 6** The set  $S^*$  is called the convex hull of the set  $S$

In the sequel the following assertion will be essential

**Theorem 2.1.1 (Caratheodory's Theorem).** *Each point  $s^*$  in the convex hull  $S^*$  of any subset  $S$ , of the  $n$ -dimensional space, can be represented in the form*

$$s^* = \sum_{i=1}^{n+1} \alpha_i s_i, \quad (2.1.2)$$

where

$$\alpha_i \geq 0, \quad \sum_{i=1}^{n+1} \alpha_i = 1, \quad s_i \in S$$

If  $s^*$  is a boundary point of the set  $S^*$ , then  $\alpha_{n+1}$  can be set equal to zero

We now prove the theorem on properties of information matrices

### Theorem 2.1.2

(1) For any design  $\epsilon$  the information matrix  $M(\epsilon)$  is a symmetric positive semi-definite matrix

(2) The matrix  $M(\epsilon)$  is degenerate ( $|M(\epsilon)| = 0$ ), if the spectrum of the design  $\epsilon$  contains less than  $m$  points ( $m$  is the number of unknown parameters)

(3) The family of matrices  $M(\epsilon)$ , corresponding to all possible normalized designs, is convex

If the function  $f(x)$  and the efficiency function  $\lambda(x)$  are continuous in the region  $X$  of possible measurements, and  $X$  is closed, then the set of information matrices is closed

(4) For any design  $\epsilon$  the matrix  $M(\epsilon)$  can be represented in the form

$$M(\epsilon) = \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i) \quad (2.1.3)$$

where

$$n \leq [(m+1)m/2] + 1, \quad 0 \leq p_i \leq 1, \quad \sum_{i=1}^n p_i = 1 \quad (2.1.4)$$

*Proof.* (1) By Definition 12 of Section 1.1, it is sufficient to show that the matrix  $M$  is symmetric and that for any vector  $z$  with real components, the quadratic form  $z' M(\epsilon) z \geq 0$ . The symmetry of the informa-

tion matrix follows from its definition, and the nonnegativeness of the corresponding quadratic form is easy to verify:

$$z' M(\epsilon) z = \int_X \lambda(x) z' f(x) f'(x) z d\epsilon(x) = \int_X \lambda(x) [z' f(x)]^2 d\epsilon(x) \geq 0. \quad (2.1.5)$$

In this way the information matrix is positive semidefinite.

(2) If the number of points in the design is finite, then by definition

$$M(\epsilon) = \sum_{i=1}^n p_i f(x_i) f'(x_i). \quad (2.1.6)$$

The rank of the matrix of the type (2.1.6) is  $r \leq n$ ; it follows that  $|M(\epsilon)| = 0$  if  $n < m$  (the dimension of the information matrix is  $m \times m$ ) and the matrix  $M(\epsilon)$  is degenerate.

(3) Let  $\epsilon_1$  and  $\epsilon_2$  be two arbitrary normalized designs given on the closed set  $X$ . Let each of these designs be characterized by the corresponding measure  $\xi_1(x)$  and the measure  $\xi_2(x)$ . Then by the linear combination of these designs  $\epsilon = \alpha \epsilon_1 + (1 - \alpha) \epsilon_2$  is understood the normalized design with measure  $\xi(x) = \alpha \xi_1(x) + (1 - \alpha) \xi_2(x)$ . It is easy to verify, relying directly on the definition of the information matrix [cf., for example, (1.10.14)], that the matrix

$$M = \alpha M(\epsilon_1) + (1 - \alpha) M(\epsilon_2)$$

is the information matrix of the normalized design  $\epsilon = \alpha \epsilon_1 + (1 - \alpha) \epsilon_2$  and is given on  $X$ , that is, belongs to the set under consideration. From this and the definition of a convex set (cf. Definition 5), it follows that the set of information matrices corresponding to continuous normalized designs, defined on  $X$ , is convex.

The closure of the set of information matrices follows from the closure of  $X$  and the continuity of  $f(x)$ .

(4) Since any information matrix is symmetric ( $M_{\alpha\beta} = M_{\beta\alpha}$ ,  $\alpha, \beta = 1, 2, \dots, m$ ), it is completely described by  $m(m + 1)/2$  elements. In other words, to each information matrix it is possible to correspond a vector of dimension  $m(m + 1)/2$ . Directly from the definition of the information matrix, it follows that the set of vectors defining the information matrices  $M(\epsilon)$ , where  $\epsilon$  is an arbitrary normalized design, is the convex closure of the set consisting of the vectors corresponding to the information matrices  $M[\epsilon(x)]$ , where the spectrum of the designs  $\epsilon(x)$  consists of single points  $x$ . From this and from Carathéodory's theorem, assertion (4) follows, and the theorem is proved.

Property (4) is particularly important from the practical viewpoint. It says that for any experimental design with number of points exceeding  $N = [(m+1)m/2] + 1$  and information matrix  $M(\epsilon)$  it is always possible to find a design  $\epsilon$  with the number of points less than or equal to  $N$ , which for this allocation (general number of distinct observations) will have the information matrix  $M(\epsilon) = M(\epsilon)$ . In other words, for the design  $\epsilon$  it is always possible to find a design with a smaller number of points but which is equivalent to it in the sense (1.9.3)-(1.9.8). In this way a distribution of observations at more than  $N = [(m+1)m/2] + 1$  points yields no advantage from the mathematical statistical point of view.

## 2.2 Equivalence of D Optimal and Minimax Designs

### Basic Properties of These Designs

I Excluding trivial cases, it is evident that universal optimal designs which simultaneously satisfy all criteria of optimality enumerated in Section 1.8 do not exist (see also the examples of that section). However, some of the criteria of optimality are strictly related to one another, and for them it is possible to construct unique optimal designs. In this section the relation between *D* optimal designs and minimax designs will be considered.

Minimax and *D* criteria compare the results of an experiment in different spaces. Therefore, the existence of designs optimal with respect to either of these criteria will always be useful for the experimenter. For  $p_i = n_i N^{-1}$  where  $n_i$  and  $N$  are integers one cannot obtain any general conclusion along these lines since the functionals (1.8.3) and (1.8.10) essentially depend on the analytic form of  $f(x)$  and on  $N$ .

In passing to continuous designs which are good approximations to reality for large  $N$ , among the *D* optimal and minimax designs it is possible to exhibit a strict relation. Indeed, the continuous *D* optimal and minimax designs are equivalent for  $\lambda(x) \equiv 1$ . In other words a *D* optimal design is also minimax, and on the other hand a minimax design is *D* optimal (Kiefer and Wolfowitz [19]).

In order to prove this fundamental assertion we must have the following auxiliary assertion.

**Lemma 2.2.1** *Let the information matrix of an arbitrary design  $\epsilon$  be continuous, then*

(1) The weighted sum of the variances of the estimates of the response surface  $d(x, \epsilon)$ , taken over all points of the design  $\epsilon$ , is equal to the number of unknown parameters  $m$ :

$$\sum_{i=1}^n p_i \lambda(x_i) d(x_i, \epsilon) = m; \quad (2.2.1)$$

in the more general case,

$$\int_X \lambda(x) d(x, \epsilon) d\xi(x) = m. \quad (2.2.2)$$

(2) The minimal value  $\max_{x \in X} \lambda(x) d(x, \epsilon)$  cannot be less than  $m$ :

$$\max_{x \in X} \lambda(x) d(x, \epsilon) \geq m. \quad (2.2.3)$$

*Proof.* (1) Using the manifest form for the dispersion of the predicted value of the response surface,

$$\begin{aligned} \int_X \lambda(x) d(x, \epsilon) d\xi(x) &= \int_X \lambda(x) f'(x) M^{-1}(\epsilon) f(x) d\xi(x) \\ &= \text{Tr } M^{-1}(\epsilon) \int_X \lambda(x) f'(x) f(x) d\xi(x) \\ &= \text{Tr } M^{-1}(\epsilon) M(\epsilon) = \text{Tr } I_m = m. \end{aligned} \quad (2.2.4)$$

In (2.24) we made use of the fact that  $\text{Tr } AB = \text{Tr } BA$  and the manifest expression for the information matrix.

(2) It is not difficult to see that

$$\int_X \lambda(x) d(x, \epsilon) d\xi(x) = d^* \int_X d\xi(x) = d^* = m, \quad (2.2.5)$$

where

$$\min_x \lambda(x) d(x, \epsilon) \leq d^* \leq \max_x \lambda(x) d(x, \epsilon). \quad (2.2.6)$$

Therefore,  $\max_x \lambda(x) d(x, \epsilon) \geq m$ . The lemma is proved.

Later, we will need the concept of a convex (concave) function.

**DEFINITION 1.** If  $X$  is a convex set, then a numerical function defined

on  $S$  is called convex if for  $s_1, s_2 \in S$  and all  $\alpha$  satisfying the condition  $0 \leq \alpha \leq 1$ ,

$$f[\alpha s_1 + (1 - \alpha) s_2] \leq \alpha f(s_1) + (1 - \alpha) f(s_2), \quad (2.2.7)$$

and is called concave if

$$f[\alpha s_1 + (1 - \alpha) s_2] \geq \alpha f(s_1) + (1 - \alpha) f(s_2) \quad (2.2.8)$$

If these inequalities are strict for  $s_1 \neq s_2, 0 < \alpha < 1$ , then the function  $f$  is called, respectively, strictly convex or strictly concave. In Fig. 3A an example of a convex function is presented. Figure 3B, shows an example of a concave function.

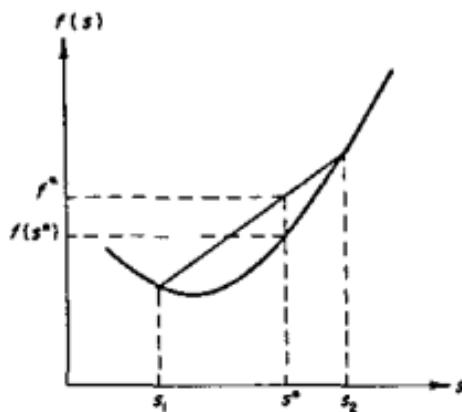
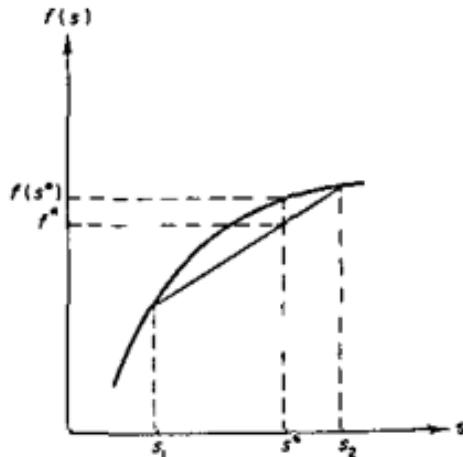


Fig. 3A Example of a convex function  $s^* = \alpha s_1 + (1 - \alpha) s_2$ ,  $f^* = \alpha f(s_1) + (1 - \alpha) f(s_2)$

Fig. 3B Example of concave function. The notation is the same as in Fig. 3A



**Lemma 2.2.2.** *The function  $\log |M(\epsilon)|$ , where  $|M(\epsilon)|$  is the information matrix of the design  $\epsilon$ , is a strictly concave function.*

*Proof.* In Theorem 2.1.2 it was shown that the set of matrices  $M(\epsilon)$  is a convex set. Therefore, to prove the lemma it is sufficient to show that

$$\log |M| > (1 - \alpha) \log |M_1| + \alpha \log |M_2|, \quad (2.2.9)$$

where  $M_1 \neq M_2$  and  $M = (1 - \alpha)M_1 + \alpha M_2$ ,  $0 < \alpha < 1$ . The inequality (2.2.9) immediately follows from the inequality (cf. Theorem 1.1.14)

$$|M| > |M_1|^{1-\alpha} |M_2|^\alpha. \quad (2.2.10)$$

The lemma is proved.

**Lemma 2.2.3.** *Let there be two designs  $\epsilon_1$  and  $\epsilon_2$  with information matrices  $M(\epsilon_1)$  and  $M(\epsilon_2)$ . Then*

$$(d/d\alpha) \log |M(\epsilon)| = \text{Tr } M^{-1}(\epsilon)[M(\epsilon_2) - M(\epsilon_1)], \quad (2.2.11)$$

where  $M(\epsilon)$  is the information matrix of the design

$$\epsilon = (1 - \alpha)\epsilon_1 + \alpha\epsilon_2, \quad 0 < \alpha < 1.$$

*Proof.* By Theorem 2.1.2,  $M(\epsilon) = (1 - \alpha)M(\epsilon_1) + \alpha M(\epsilon_2)$ . Differentiating  $\log |(1 - \alpha)M(\epsilon_1) + \alpha M(\epsilon_2)|$  with respect to  $\alpha$ , we obtain

$$\frac{d}{d\alpha} \log |M(\epsilon)| = \text{Tr } M^{-1}(\epsilon) \frac{d}{d\alpha} M(\epsilon) = \text{Tr } M^{-1}(\epsilon)[M(\epsilon_1) - M(\epsilon_2)]. \quad (2.2.12)$$

In (2.2.12) we used the results of Part VI, Section 1.1. The lemma is proved.

**Theorem 2.2.1 (Equivalence Theorem [19]).** *The following assertions:*

- (1) *the design  $\hat{\epsilon}$  maximizes  $|M(\epsilon)|$  [minimizes  $|D(\epsilon)|$ ],*
- (2) *the design  $\hat{\epsilon}$  minimizes  $\max_x \lambda(x) d(x, \epsilon)$ ,*
- (3)  $\max_x \lambda(x) d(x, \hat{\epsilon}) = m$

*are equivalent. The information matrices of all designs satisfying (1)–(3) coincide among themselves. Any linear combination of designs satisfying (1)–(3) also satisfies (1)–(3).*

*Proof* (1) We will show that (2) follows from (1). Let the design  $\hat{\epsilon}$  maximize  $|M(\epsilon)|$ .

We consider the design corresponding to the linear combination of the design  $\hat{\epsilon}$  and some arbitrary design  $\epsilon$ :  $\tilde{\epsilon} = (1 - \alpha)\hat{\epsilon} + \alpha\epsilon$ . By Lemma 2.2.3,

$$\begin{aligned}(d/dx) \log |M(\epsilon)||_{\alpha=0} &= \text{Tr } M^{-1}(\hat{\epsilon})[M(\epsilon) - M(\hat{\epsilon})]|_{\alpha=0} \\ &= \text{Tr } M^{-1}(\hat{\epsilon})M(\epsilon) - m\end{aligned}$$

In view of the definition of the design  $\hat{\epsilon}$  the value of the derivative must be less than or equal to zero.

Assuming that the spectrum of the design  $\epsilon$  consists of one point  $x$ , belonging to  $X$ , we obtain

$$\begin{aligned}\text{Tr } M^{-1}(\hat{\epsilon})M[\epsilon(x)] - m &= \text{Tr } M^{-1}(\hat{\epsilon})\lambda(x)f(x)f'(x) - m \\ &= \lambda(x)d(x, \hat{\epsilon}) - m < 0\end{aligned}\quad (2.2.13)$$

On the other hand, from Lemma 2.2.2 for any design,

$$\max_x \lambda(x)d(x, \hat{\epsilon}) \geq m \quad (2.2.14)$$

Comparison of (2.2.13) and (2.2.14) shows that the  $D$  optimal design minimizes  $\max_x \lambda(x)d(x, \hat{\epsilon})$ .

(2) Let the design  $\hat{\epsilon}$  minimizing  $\max_x \lambda(x)d(x, \hat{\epsilon})$ , not be  $D$ -optimal. Then by Lemma 2.2.2 there is a design  $\epsilon$  such that

$$(d/dx) \log |(1 - \alpha)M(\hat{\epsilon}) + \alpha M(\epsilon)||_{\alpha=0} = \text{Tr } M^{-1}(\hat{\epsilon})M(\epsilon) - m > 0 \quad (2.2.15)$$

By Theorem 2.1.2, any design  $\epsilon$  can be represented in the form of a superposition of  $[m(m+1)/2] + 1$  designs  $M[\epsilon(x_i)]$  ( $i = 1, 2, \dots, [m(m+1)/2] + 1$ ). Therefore, without loss of generality, we may consider that the design  $\epsilon$  consists of a finite number of points. Then

$$\text{Tr } M^{-1}(\hat{\epsilon})M(\epsilon) - m = \sum_{i=1}^n p_i \lambda(x_i)d(x_i, \hat{\epsilon}) - m \quad (2.2.16)$$

But the design  $\hat{\epsilon}$  is maximal, that is,  $\lambda(x)d(x, \hat{\epsilon}) \leq m$  [cf. (1) of proof] and it follows that

$$\sum_{i=1}^n p_i \lambda(x_i)d(x_i, \hat{\epsilon}) - m \leq m \sum_{i=1}^n p_i - m = 0 \quad (2.2.17)$$

The inequalities (2.2.15) and (2.2.17) coincide only if the design minimizing  $\max_x \lambda(x) d(x, \epsilon)$  is  $D$ -optimal.

(3) The equivalence of assertions (1) and (3) and (2) and (3) immediately follows from the equivalence of assertions (1) and (2) and Lemma 2.2.1.

(4) Let the designs  $\epsilon_1$  and  $\epsilon_2$  with information matrices  $M(\epsilon_1)$  and  $M(\epsilon_2)$  be  $D$ -optimal and  $M(\epsilon_1) \neq M(\epsilon_2)$ .

We consider the information matrix corresponding to the composition of designs  $\epsilon_1$  and  $\epsilon_2$ :  $M(\epsilon) = (1 - \alpha)M(\epsilon_1) + \alpha M(\epsilon_2)$ . Corresponding to Lemma 2.2.2,

$$\log |M(\epsilon)| > (1 - \alpha) \log |M(\epsilon_1)| + \alpha \log |M(\epsilon_2)|. \quad (2.2.18)$$

But according to the definition of a  $D$ -optimal design

$$|M(\epsilon_1)| = |M(\epsilon_2)| \geq |M(\epsilon)|. \quad (2.2.19)$$

It is not difficult to see that (2.2.18) and (2.2.19) do not contradict one another only if

$$M(\epsilon_1) = M(\epsilon_2) = M(\epsilon). \quad (2.2.20)$$

Considering the results (1)–(3) and (2.2.20), it is not difficult to obtain the validity of the concluding part of the theorem. The theorem is proved.

Theorem 2.2.1 plays an important role in constructing the mathematical apparatus of designs. From it, in particular, it follows that for  $\lambda(x) = 1$  the minimax [in the sense (1.9.7)] continuous designs are equivalent to continuous  $D$ -optimal designs. Therefore, having proved any property for the information matrix of a  $D$ -optimal design, we can be sure of the validity of this property for the information matrix of the minimax design. On the other hand, in constructing optimal designs one can interchangeably use the properties of  $D$ -optimal and minimax designs. As we will see below, this permits us in several cases to find optimal designs by simple and elementary methods.

Theorem 2.2.1 also gives a very simple method for verifying  $D$ -optimality of a design. For this it is sufficient to verify that the dispersion  $\lambda(x) d(x, \epsilon)$  does not exceed  $m$ . In this case it is very useful to obtain the following corollary of Theorem 2.2.1.

**Corollary 1.** At the points of the optimal design  $\epsilon$  the dispersion  $d(x, \epsilon)$  attains its maximum value  $m$

We assume the contrary

$$\lambda(x) d(x, \epsilon) < m, \quad (2.2.21)$$

where  $x$  is one of the points of the design  $\epsilon$ . Then, in view of (3) of Theorem 2.2.1,

$$\sum_{i=1}^n p_i \lambda(x_i) d(x_i, \epsilon) < \sum_{i=1}^n p_i m = m$$

But by Lemma 2.2.1

$$\sum_{i=1}^n p_i \lambda(x_i) d(x_i, \epsilon) = m$$

The contradiction obtained proves our assertion

Corollary 1 is particularly useful in that instead of verifying that the inequality

$$\lambda(x) d(x, \epsilon) \leq m$$

is satisfied over the entire region  $X$ , we verify only that the equality

$$\lambda(x) d(x, \epsilon) = m \quad (2.2.22)$$

is satisfied for points  $x_i$  ( $i = 1, 2, \dots, m$ ) of the design

We note that equality (2.2.22) satisfied at points of the  $D$ -optimal design is a necessary condition but is not sufficient

We formulate another simple but very useful assertion for constructing  $D$ -optimal designs

**Theorem 2.2.2** If  $\epsilon_1$  and  $\epsilon_2$  are two designs with distinct information matrices  $M(\epsilon_1)$  and  $M(\epsilon_2)$ , for which  $|M(\epsilon_1)| = |M(\epsilon_2)|$ , then the design

$$\epsilon = (1 - \alpha)\epsilon_1 + \alpha\epsilon_2, \quad 0 < \alpha < 1$$

has the determinant

$$|M(\epsilon)| > |M(\epsilon_1)|$$

The proof of the given theorem follows immediately from the strict concavity of  $\log |M(\epsilon)|$

The equivalence theorem 2.2.1 and Theorem 2.2.2 permit the construction of  $D$ -optimal designs, in many not very complicated regression problems, by skipping complicated computations and relying basically on semi-intuitive considerations.

Usually in this procedure the search for optimal designs consists of the following. From symmetry considerations or from the analytic form of the response surface, the set  $X$  is divided into families of points at which the measurements are allocated in the obvious way. Inside of this family the minimization of  $|D(\epsilon)|$  or the minimization of  $\max_x \lambda(x) d(x, \epsilon)$  in  $\epsilon$  is carried out, and it is verified afterwards that they satisfy either of the design conditions of Theorem 2.2.1 (or its Corollary 1).

EXAMPLE 1. Let

$$E(y | x) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \cdots + \theta_m x_{m-1}.$$

The region  $X$ , where the measurements are taken, is a hypersurface with center at the origin and radius equal to unity:  $x'x \leq 1$ . Since the dispersion matrix  $D(\epsilon)$  of the parameters  $\theta_1, \theta_2, \dots, \theta_m$  for any nondegenerate design is positive definite, the quadratic form  $d(x, \epsilon) = f'(x) D(\epsilon) f(x)$ , where  $f'(x) = \|1, x_1, x_2, \dots, x_{m-1}\|$ , attains its maximal value on the boundary of the region  $X$  (in this case the hypersphere with radius equal to unity). It follows that the points of the  $D$ -optimal design must lie on the given surface.

It is well known that the levels of constant values of a positive-definite quadratic form are ellipsoids. But, as is easy to see, any ellipsoid described around the hypersphere and touching it in  $n \geq m$  points not lying in a single hyperplane coincides with this hypersphere. That is, for the  $D$ -optimal design  $\hat{\epsilon}$ , the quadratic form  $d(x, \hat{\epsilon})$  must have hyperspheres as surfaces of constant values. This is equivalent to the matrix  $D(\hat{\epsilon})$  coinciding with the diagonal matrix whose elements  $D_{\alpha\alpha}(\hat{\epsilon})$  ( $\alpha = 1, 2, \dots, m$ ) are equal to one another.

If we take into account that  $D_{11}(\epsilon) = 1$  (since  $M_{11} = \sum_{i=1}^n p_i = 1$ ), then the unique matrix, satisfying the supplementary conditions (cf. Corollary 1 of Theorem 2.2.1)

$$d(x, \hat{\epsilon})|_{x'x=1} = f'(x) D(\hat{\epsilon}) f(x) |_{x'x=1} = \sum_{\alpha=1}^m D_{\alpha\alpha}(\hat{\epsilon}) = m,$$

is the identity matrix  $I_m$ . Therefore the equations

$$M_{(k)(k)}(\xi) = \sum_{i=1}^n p_i x_{(k-1)i} x_{(k-1)i} = 1,$$

$$M_{(k-1)(j-1)}(\xi) = \sum_{i=1}^n p_i x_{ki} x_{ji} = 0,$$

$$M_{(k-1)1}(\xi) = \sum_{i=1}^n p_i x_{ki} = 0$$

must be satisfied. From geometric considerations, it is not difficult to see that the given conditions are satisfied by designs whose spectrum coincides with the vertices of any regular polygon, inscribed in the hypersphere.

**EXAMPLE 2 [20]** Let the response surface be a surface of the second order and the region  $X$  coincide with the hypercube

$$F(y | x) = \sum_{s=1}^m \theta_s f_s(x)$$

$$\lambda(x) = 1, \quad -1 \leq x_k \leq 1 \quad k = 1, 2, \dots, q \quad m = (q+1)(q+2)/2$$

The admissible functions  $f_s(x)$  can be defined in the following way

$$f_1(x) = 1, \quad f_{1+j}(x) = x_j^2, \quad 1 \leq j \leq q,$$

$$f_{q+1+r}(x) = x_r, \quad 1 \leq r < q$$

$f_j(x)$  for  $2q + 2q + 2 \leq j \leq (q+1)(q+2)/2$  are the functions  $x_p x_r$ ,

$p < r$  in any order

It is easy to verify that the dimension of the vector  $f(x)$  is equal to  $m = (q+1)(q+2)/2$ .

From symmetry considerations it is natural to assume that one of the  $D$ -optimal designs belongs to the set of designs having the form. Observations with weights  $\alpha$  are taken at each of the  $2^q$  vertices of the  $q$ -dimensional cube, with weights  $\beta$  at each of the  $q2^{q-1}$  points which are in the middle of the edges, and with weights  $\gamma$  at each of the  $q(q-1)2^{q-3}$  centers of the two-dimensional faces. The spectrum of the designs consists of  $n = 2^{q-3}[8 + 4q + q(q-1)]$  points.

After  $\min |D|$  with respect to  $\alpha, \beta, \gamma (\sum \alpha + \sum \beta + \sum \gamma = 1)$  is found, we verify that the inequality  $\max_x d(x) \leq m$  is satisfied.

If this inequality is satisfied, then by Theorem 2.2.1 we can find a design which will be  $D$ -optimal.

It is easy to verify that the information matrix constructed according to the assumed design will have the form

$$M(\epsilon) = \begin{vmatrix} 1 & F & 0 & 0 \\ F' & G & 0 & 0 \\ 0 & 0 & uI_q & 0 \\ 0 & 0 & 0 & vI_{q(q-2)/2} \end{vmatrix}.$$

Here

$$\begin{aligned} u &= \sum_{i=1}^n x_{1i}^2 p_i = 2^{q-3}[8\alpha + 4(q-1)\beta + (q-1)(q-2)\gamma], \\ v &= \sum_{i=1}^n x_{1i}^2 x_{2i}^2 p_i = 2^{q-3}[8\alpha + 4(q-2)\beta + (q-2)(q-3)\gamma], \end{aligned} \quad (2.2.23)$$

where  $p_i$  is equal to one of the three values  $\alpha, \beta, \gamma$ . All of the elements of the column vector  $F$  are equal to  $F_j = u$  ( $j = 1, 2, \dots, q$ ). The matrix  $G$  has dimension  $q \times q$ ; its diagonal elements are equal to  $u$ , and the nondiagonal elements are equal to  $v$ . The symbol 0 indicates the null matrix. The determinant of the matrix  $M(\epsilon)$  is equal to

$$|M(\epsilon)| = u^q v^{q(q-1)/2} (u - v)^{q-1} [u + (q-1)v - qu^r].$$

Its inverse matrix is equal to

$$D(\epsilon) = \begin{vmatrix} A & B & 0 & 0 \\ B' & C & 0 & 0 \\ 0 & 0 & u^{-1}I_q & 0 \\ 0 & 0 & 0 & v^{-1}I_{q(q-1)/2} \end{vmatrix}.$$

Solving the system of equations

$$\partial |M| / \partial u = 0, \quad \partial |M| / \partial v = 0$$

under the condition  $\alpha > 0, \beta > 0, \gamma > 0$ , we obtain

$$u = \frac{q+3}{4(q+1)(q+2)^2} [2q^2 + 3q + 7 + (q-1)(4q^2 + 12q + 17)^{1/2}],$$

$$v = \frac{q+3}{8(q+1)(q+2)^3} [4q^3 + 8q^2 + 11q - 5 + (2q^2 + q + 3)(4q^2 + 12q + 17)^{1/2}].$$

For given  $u$  and  $v$ , after a straightforward but cumbersome computation, we obtain

$$d(x, \epsilon) = [(q+1)(q+2)/2] - c \sum_{i=1}^m (x_i^2 - x_i^4),$$

where  $c$  is a positive constant. It is not difficult to see that in the region  $-1 \leq x_k \leq 1$ ,  $k = 1, 2, \dots, q$

$$\max_x d(x, \epsilon) = (q+1)(q+2)/2$$

From system (2.2.23) and the condition of normalized weights

$$\sum \alpha + \sum \beta + \sum \gamma = 2^{q-3}[8\alpha + 4q\beta + q(q-1)\gamma] = 1,$$

we can find the optimal  $\alpha, \beta, \gamma$

$$\alpha = 2^{-q-1}[(q-1)(q-2) - 2q(q-2)u + q(q-1)v]$$

$$\beta = 2^{-q+3}[(2q-3)u - (q-1)v - (q-2)],$$

$$\gamma = 2^{2-q}[1 + v - 2u]$$

The numerical values for  $\alpha, \beta$ , and  $\gamma$  for  $q \leq 5$  are presented in Table 1

Table 1  
The Numerical Values for  $\alpha, \beta$  and  $\gamma$  for  $q \leq 5$

	$\alpha$	$\beta$	$\gamma$
1	0.333	0.333	0.000
2	0.1458	0.08015	0.0962
3	0.071975	0.01895	0.03280
4	0.03705	0.0038375	0.01185
5	0.01928	0.0003125	0.004475

For  $q \geq 6$  the above method for the construction of  $D$ -optimal designs is not suitable, since  $\beta < 0$  ( $q \geq 6$ ), which is impossible.

II. For the continuous optimal designs under consideration, in view of Theorem 2.2.1 for  $\lambda(x) = 1$ ,  $x \in X$ , it is not necessary to distinguish between  $D$ -optimal and minimax designs. Therefore,

in the future we will speak only of  $D$ -optimal continuous designs understanding that all of the results obtained will be equally valid for minimax continuous designs.

We will show that for  $D$ -optimal designs an upper bound of the minimal number of points turns out to be one less than for an arbitrary design.

**Theorem 2.2.3.** *If the set  $X$  is closed and the functions  $\lambda(x), f(x)$  are continuous, then for any  $D$ -optimal design  $\hat{\epsilon}$  with information matrix  $M(\hat{\epsilon})$  and number of points  $n > m(m + 1)/2$ , a design  $\hat{\epsilon}'$  can always be found with number of points  $N \leq m(m + 1)/2$ , and the determinant of the information matrix satisfies*

$$|M(\hat{\epsilon})| = |M(\hat{\epsilon}')|.$$

*Proof.* Since the collection of information matrices is the convex hull of the information matrices  $M[\epsilon(x)]$ , corresponding to the point designs  $\epsilon(x)$ , it follows from Carathéodory's theorem that any boundary point of this set can be expressed in the form of a linear combination

$$M(\epsilon) = \sum_{i=1}^q p_i M[\epsilon(x_i)], \quad q = m(m + 1)/2,$$

where  $p_i \geq 0$ ,  $\sum_{i=1}^q p_i = 1$  and  $M[\epsilon(x_i)]$  are the information matrices of the designs concentrated at the points  $x_i$ .

Therefore, for the proof of the theorem it is sufficient to show that the information matrices corresponding to  $D$ -optimal designs are boundary points of the set of all information matrices.

We assume that  $M(\hat{\epsilon})$  is an interior point of the given set. Then there exists a positive number  $\alpha$  such that the matrix  $M(\epsilon) = (1 + \alpha) M(\hat{\epsilon})$  also belongs to the set of information matrices under consideration, i.e.,  $M(\epsilon)$  is the information matrix of some design  $\epsilon$ . But

$$|(1 + \alpha) M(\hat{\epsilon})| = (1 + \alpha)^m |M(\hat{\epsilon})| > |M(\hat{\epsilon})|,$$

which contradicts the definition of the optimal design. From this, the matrix  $M(\hat{\epsilon})$  is a boundary point, which proves the theorem.

Theorem 2.2.3 is particularly useful for computational methods of constructing  $D$ -optimal designs, since it permits us to reduce the dimension of the extremal problem.

III. It is usually desirable that the optimal designs possess a possibly large number of invariance properties. For example, for the experimenter it is very useful when the optimal design is invariant with respect to passage of one criterion of optimality to another. As an example of such invariance we may consider the equivalence of continuous  $D$ -optimal and minimax designs [ $\lambda(x) = 1$ ]. The prospect of constructing plans which would be optimal not only for the given admissible function  $f'(x) = \|f_1(x), f_2(x), \dots, f_n(x)\|$ , but for a possibly wider family of admissible functions  $f'(x) = \|f_1(x), f_2(x), \dots, f_m(x)\|$  is extraordinarily appealing. We investigate the degree to which  $D$ -optimal designs satisfy the stated properties of invariance.

**Theorem 2.2.4.** *Let  $L$  be some nondegenerate ( $|L| \neq 0$ ) linear transformation*

$$\varphi(x) = If(x), \quad (2.2.24)$$

*then, if the design  $\hat{\epsilon}$  is  $D$ -optimal with respect to the admissible function  $f(x)$ , it is  $D$ -optimal with respect to the admissible function  $\varphi(x)$ .*

*Proof.* For some design  $\epsilon$ , let the dispersion matrix have the value  $D(\epsilon)$  for  $\eta(x, \theta) = \theta f(x)$ . The regression function  $\eta(x, \theta)$  can also be written in the form

$$\eta(x, \theta) = \rho \varphi(x) = \rho L^{-1} f(x) \quad (2.2.25)$$

Since (2.2.25) holds for any  $x$ ,  $\rho = L' \theta$ . Corresponding to Corollary 4 of Theorem 1.3.1  $|D(\hat{\theta}, \epsilon)| = |L'|^2 |D(\theta, \epsilon)| L$  or

$$|D(\hat{\theta}, \epsilon)| = |L|^2 |D(\theta, \epsilon)| \quad (2.2.26)$$

Minimizing both sides of (2.2.26) with respect to  $\epsilon$  we obtain

$$\min_{\epsilon} |D(\hat{\theta}, \epsilon)| = |L|^2 \min_{\epsilon} |D(\theta, \epsilon)| = |D(\hat{\theta}, \hat{\epsilon})| = |L|^2 |D(\theta, \hat{\epsilon})|,$$

that is,  $\min_{\epsilon} |D(\hat{\theta}, \epsilon)| = |D(\hat{\theta}, \hat{\epsilon})|$ , which was what was required to prove.

Sometimes Theorem 2.2.4 is more useful to use in a slightly different form

**Theorem 2.2.4a.**  *$D$ -optimal designs are invariant with respect to any non degenerate linear transformation of estimated parameters:*

$$\rho = C\theta, \quad |C| \neq 0. \quad (2.2.27)$$

In order to pass from (2.2.27) to (2.2.24) it is sufficient to set  $C = L'$ .

In Theorem 2.2.4 it was shown that  $D$ -optimal designs are invariant with respect to the transformation (2.2.24). There exist transformations with respect to which  $D$ -optimal designs are not invariant, and are transformed according to simple rules.

**Theorem 2.2.5.** *Let the control variables  $x$  be transformed according to the rule*

$$z = Z(x), \quad (2.2.28)$$

where  $Z$  is a nondegenerate single-valued transformation in the region  $X$ :

$$\left| \frac{D(z)}{D(x)} \right| = \begin{vmatrix} \frac{\partial Z_1(x)}{\partial x_1} & \frac{\partial Z_1(x)}{\partial x_2} & \dots & \frac{\partial Z_1(x)}{\partial x_k} \\ \frac{\partial Z_2(x)}{\partial x_1} & \frac{\partial Z_2(x)}{\partial x_2} & \dots & \frac{\partial Z_2(x)}{\partial x_k} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial Z_k(x)}{\partial x_1} & \frac{\partial Z_k(x)}{\partial x_2} & \dots & \frac{\partial Z_k(x)}{\partial x_k} \end{vmatrix} \neq 0,$$

and let  $\hat{\epsilon}_x$  be a  $D$ -optimal design for the regression function  $\eta(x, \theta) = \theta'f(x)$  with distribution function of the allocation  $p(x)$  defined on  $X$ .

Then the distribution function of the allocation for the  $D$ -optimal design  $\hat{\epsilon}_z$  of the regression function  $\eta(z, \theta) = \theta'\varphi(z)$  is equal to

$$p(z) = \left| \frac{D(z)}{D(x)} \right| p(x). \quad (2.2.29)$$

For the discrete spectrum, the design  $\hat{\epsilon}_z$  is related to the design  $\hat{\epsilon}_x$  in the following way:

$$\hat{\epsilon}_z = \left\{ \begin{array}{l} z_1 = Z(x_1), z_2 = Z(x_2), \dots, z_n = Z(x_n) \\ p_{z1} = p_{x1}, p_{z2} = p_{x2}, \dots, p_{zn} = p_{xn} \end{array} \right\}. \quad (2.2.30)$$

*Proof.* For the proof of the theorem it is sufficient to verify that

$\lambda(z) d(z, \hat{\epsilon}_s) \leq m$  for  $z = Z(x)$ ,  $x \in X$ . Indeed, for the design  $\hat{\epsilon}_s$ , the determinant (2.2.29), we have

$$\begin{aligned}\lambda(z) d(z, \hat{\epsilon}_s) &= \lambda(z) \varphi'(z) \left[ \int_z \lambda(\tau) p(\tau) \varphi(\tau) \varphi'(\tau) d\tau \right]^{-1} \varphi(z) \\&= \lambda[Z(x)] \varphi'[Z(x)] \left\{ \int_x \lambda[Z(\tau)] p(\tau) \left| \frac{D(\tau)}{D(z)} \right| \right. \\&\quad \times \varphi[Z(\tau)] \varphi'[Z(\tau)] \left| \frac{D(\tau)}{D(x)} \right| d\tau \left. \right\}^{-1} \varphi[Z(x)] \\&= \lambda(x) f'(x) \left[ \int_x \lambda(\tau) p(\tau) f(\tau) f'(\tau) d\tau \right]^{-1} f(x) \\&= \lambda(x) d(x, \hat{\epsilon}_s),\end{aligned}$$

since  $\lambda[Z(x)] = \lambda(x)$  and  $\varphi[Z(x)] = f(x)$ . Therefore,

$$\max_{z \in Z} \lambda(z) d(z, \hat{\epsilon}_s) = \max_{x \in X} \lambda(x) d(x, \hat{\epsilon}_s) \leq m \quad (2.2.31)$$

From (2.2.31) and the equivalence theorem 2.2.1 it follows that  $p(z)$ , defined by (2.2.29), corresponds to the  $D$ -optimal design in the space  $Z$ .

The discrete case is considered analogously (integration in this case is replaced by summation).

The theorem presented permits us in many cases to transform the control variables of the functions  $\eta(x, \theta)$  and  $\lambda(x)$  to corresponding functions for which  $D$ -optimal designs are known or can be comparatively easily constructed.

**EXAMPLE 3** Let

$$\eta(x, \theta) = \theta_1 + \theta_2 e^{-\lambda_1 x},$$

$$\lambda(x) = e^{-\lambda_1 x}, \quad 0 \leq x \leq \infty$$

It is required to find the  $D$ -optimal design  $\hat{\epsilon}_s$ . We introduce the change of variable  $z = e^{-\lambda_1 x}$ . Then the regression function has the form  $\eta(z, \theta) = \theta_1 + \theta_2 z$ , and the efficiency function will be equal to  $\lambda(z) = z$ ,  $0 \leq z \leq 1$ . For this regression problem it is easy to find the  $D$ -optimal design.

We consider the design of the type

$$\epsilon_s = \left\{ \frac{1}{1-p}, \frac{z}{p} \right\} \quad 0 \leq z \leq 1, \quad 0 < p < 1.$$

For such designs the information matrix is equal to

$$M(\epsilon_z) = \begin{vmatrix} 1-p+zp & 1-p+z^2p \\ 1-p+z^2p & 1-p+z^3p \end{vmatrix},$$

and its determinant

$$|M(\epsilon_z)| = (1-p)p(z - 2z^2 + z^3).$$

The maximum of this determinant, as is easy to verify, takes place when  $p = \frac{1}{2}$  and  $z = \frac{1}{3}$ . By Theorem 2.2.1 it is not difficult to verify that for such a choice of  $p$  and  $z$  the design is  $D$ -optimal.

Passing to the old variable  $x$  and relying on Theorem 2.2.5 we obtain

$$\epsilon_x = \left\{ \begin{array}{ll} x_1 = 0, & x_2 = \lambda_1^{-1} \ln 3 \\ \frac{1}{2}, & \frac{1}{2} \end{array} \right\}.$$

### 2.3. One-Dimensional Polynomial Regression

I. We now pass to the explicit construction of  $D$ -optimal designs for concrete functions  $f_\alpha(x)$  ( $\alpha = 1, 2, \dots, m$ ). In the majority of real problems, designs with a minimal number of points at which it is necessary to take measurements are particularly useful. Therefore in constructing optimal designs it is natural to aim for constructing continuous normalized designs having a spectrum consisting of a small number of points.

From the computational viewpoint, the search for a  $D$ -optimal design consists of minimizing with respect to  $p_i$  and  $x_i$  ( $i = 1, 2, \dots, n$ ;  $\sum_{i=1}^n p_i = 1$ ) the determinant

$$|D(\epsilon)| = |M^{-1}(\epsilon)| = \left| \left[ \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i) \right]^{-1} \right|. \quad (2.3.1)$$

In many cases the extremal problem in the space of  $n(k + 1)$  variables, where  $k$  is the dimension of the factor space, reduces to several extremal problems each in a space of a small number of variables. This sharply reduces the amount of computation and sometimes results in an analytic solution. Such a simplification is possible because of the fact that the determinant  $|M(\epsilon)|$  can be represented in a special form.

**Theorem 2.3.1**

$$|M(\epsilon)| = \sum p_{j_1} p_{j_2} \dots p_{j_m} \lambda(x_{j_1}) \lambda(x_{j_2}) \dots \lambda(x_{j_m}) F^2 \begin{pmatrix} 1 & 2 & \dots & m \\ j_1 & j_2 & \dots & j_m \end{pmatrix} \quad (2.3.2)$$

where the sum is taken over all possible minor matrices of

$$F = \begin{vmatrix} f(x_1) & f(x_2) & \dots & f(x_n) \end{vmatrix}$$

*Proof* The result (2.3.2) follows immediately from Theorem 1.1.2 if we set  $B_{ii} = p_i \lambda(x_i)$  and  $A = F$ .

**EXAMPLE** Let

$$\eta(x, \theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2, \quad \lambda(x) = 1 \quad \text{and} \quad -1 \leq x_1 \leq 1, \quad -1 \leq x_2 \leq 1$$

We consider the design concentrated in three vertices of the square with equal weights  $p_i = \frac{1}{3}$  ( $i = 1, 2, 3$ ). From symmetry considerations, it is clearly possible to construct four such designs with identical determinants of the information matrices, equal to  $|M(\epsilon_3)| = \frac{1}{27} |F_3|^2$ . Here the index indicates that the spectrum of the design consists of three points and

$$F_3 = \begin{vmatrix} 1 & 1 & 1 \\ x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{vmatrix}$$

Using Theorem 2.3.1 it is easy to show that these designs are not optimal. Indeed, consider the design concentrated at all vertices of the cube with equal weights  $p_i = \frac{1}{4}$  ( $i = 1, 2, 3, 4$ ).

Corresponding to (2.3.2) the determinant of the information matrix

$$|M(\epsilon_4)| = \sum p_{j_1} p_{j_2} p_{j_3} p_{j_4} \left| F \begin{pmatrix} 1 & 2 & 3 \\ j_1 & j_2 & j_3 \end{pmatrix} \right|^2 = \sum \frac{1}{64} |F_3|^2 = \frac{1}{16} |F_3|^2$$

In this manner

$$|M(\epsilon_4)| = \frac{1}{27} |F_3|^2 > \frac{1}{27} |F_3|^2 = |M(\epsilon_3)|$$

**Corollary 1** Equation (2.3.2) takes on a particularly simple form when  $n = m$

$$|M(\epsilon)| = \prod_{i=1}^m p_i \prod_{i=1}^m \lambda(x_i) |F|^2 \quad (2.3.3)$$

If from any considerations whatever it is possible to show that the  $D$ -optimal design consists of  $m$  points then it is not difficult to see that all weights must be distributed uniformly:  $p_1 = p_2 = \dots = m^{-1}$ . Because of this the search for an optimal design will consist in minimizing, with respect to  $x_1, x_2, \dots, x_m$ , the functional

$$\prod_{i=1}^n \lambda(x_i) |F|^2. \quad (2.3.4)$$

II. We now delve into more detail for the important particular case of power regression. The functions  $f_\alpha(x) = x^{\alpha-1}$  ( $\alpha = 1, 2, \dots, m$ ) are a sufficiently coarse system of functions and provide an adequate description of any smooth function on a given interval (for more details cf., for example, [17]). For the investigation of  $D$ -optimal designs in the case of one-dimensional polynomial regression, we will use some results that are well known in the theory of approximation of functions.

**DEFINITION 1.** The system of continuous functions  $f_\alpha(x)$ ,  $\alpha = 1, 2, \dots, m$ , defined on the interval  $[a, b]$  is called a Chebyshev system on  $[a, b]$ , if any linear combination of these functions  $\sum_{\alpha=1}^m a_\alpha f_\alpha(x)$  ( $a_\alpha$  is real and  $\sum_{\alpha=1}^m a_\alpha^2 > 0$ ) has no more than  $m - 1$  roots on the given interval.

The stated conditions are equivalent to the determinant

$$\begin{vmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ f_1(x_m) & f_2(x_m) & \cdots & f_m(x_m) \end{vmatrix} \quad (2.3.5)$$

having one and the same sign for any  $a \leq x_1 < x_2 < \dots < x_m \leq b$ . For definiteness, unless stated otherwise, we will assume that the given determinant is positive.

The classic example of a Chebyshev system is the system of power functions  $f_\alpha(x) = x^{\alpha-1}$  ( $\alpha = 1, 2, \dots, m$ ).

We investigate the question of the number of points of the design for polynomial regression.

**Theorem 2.3.2.** [21, 22]. Let  $\eta(x, \theta) = \sum_{\alpha=1}^m \theta_\alpha x^{\alpha-1}$ . Then the  $D$ -optimal design is concentrated at  $m$  points if one of the following conditions is satisfied:

- (1) The system of functions  $1, \lambda(x), \lambda(x)x, \dots, \lambda(x)x^{(2m-1)}$  is a Chebyshev system of functions on the interval  $[a, b]$ .
- (2)  $\lambda(x) = P^{-1}(x)$ , where  $P(x)$  is a positive polynomial on  $[a, b]$  and its  $(2m-1)$ st derivative  $P^{(2m-1)}(x)$  has no zero on the open integral  $(a, b)$ .
- (3)  $\lambda(x) = P^{-1}(x)$ , where  $P(x)$  is a polynomial of degree no higher than  $2(m-1)$  and positive on  $[a, b]$ .

*Proof* By Theorem 2.2.1 and its Corollary 1, at the points of the  $D$ -optimal design,

$$m - \lambda(x)d(x, \hat{\epsilon}) = 0, \quad (2.3.6)$$

and at the remaining points

$$m - \lambda(x)d(x, \hat{\epsilon}) > 0 \quad (2.3.7)$$

The dispersion matrix  $D(\hat{\epsilon})$  is positive definite and it follows that

$$d(\hat{\epsilon}) = f(x)D(\hat{\epsilon})f(x) - \sum_{s=0}^{2m-2} a_s x^s \geq 0 \quad (2.3.8)$$

where

$$a_s = \sum_{s \leq t \leq 2m-2} D_{st}(\hat{\epsilon})$$

It is evident that the number of points of the design must be greater than or equal to  $m$ , otherwise the information matrix will be degenerate.

We assume that the  $D$ -optimal design consists of at least  $m+1$  points. We consider the function

$$\Psi(x) = m - \gamma - \lambda(x) \sum_{s=0}^{2m-2} a_s x^s \quad (2.3.9)$$

where  $\gamma$  is a small positive number.

If all points of the design are found inside  $[a, b]$  then by (2.3.6) and (2.3.7) the number of distinct roots of the function  $\Psi(x)$  is larger than or equal to  $2(m+1)$  (cf. Fig. 4). If one or two points of the design are at the end, then the number of distinct roots will be respectively no less than  $2m+1$  or  $2m$ . In this way, for polynomial regression of order  $m-1$  the number of distinct roots of the function  $\Psi(x)$  is

$$n \geq 2m \quad (2.3.10)$$

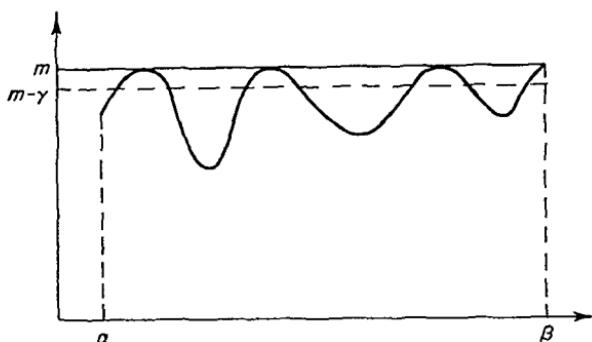


Fig. 4. Dependence of  $\Psi(x)$  on  $x$  for the optimal design.

On the other hand, the following assertions are valid:

1. If the system of functions  $\{1, \lambda(x), \lambda(x)x, \dots, \lambda(x)x^{2m-2}\} = \{\varphi_1(x), \varphi_2(x), \dots, \varphi_{2m}(x)\}$  is Chebyshev on  $[a, b]$ , then corresponding to Definition 1 any polynomial  $\sum_{i=1}^{2m} c_i \varphi_i(x)$  has on  $[a, b]$  no more than  $2m - 1$  distinct roots.

2. The polynomial

$$(m - \gamma) P(x) - \sum_{\alpha=0}^{2m-2} a_\alpha x^\alpha, \quad P(x) > 0$$

corresponding to Polya's theorem [23] and condition 2 of the theorem has no more than  $2m - 1$  roots.

3. In a given case, the function

$$(m - \gamma) P(x) - \sum_{\alpha=0}^{2m-2} a_\alpha x^\alpha, \quad P(x) > 0$$

is a polynomial of order  $2m - 2$ , and it follows that it has less than  $2m - 2$  roots.

Each of the stated assertions contradicts (2.3.10). From this it follows that the number of points of the design, in all three cases given in the conditions of the theorem, equals  $m$ .

By means of passage to the limit, it is not difficult to show that for those  $\lambda(x)$  which can be uniformly approximated by type-2 functions, the  $D$ -optimal designs also consist of  $m$  points.

There exist, however, functions of effectiveness for which  $D$ -optimal designs for the polynomial regression have more than  $m$  points.

**EXAMPLE** We consider linear one-dimensional regression  $\eta(x, \theta) = \theta_1 + \theta_2x$  on the integral  $[-1, 1]$ . The function of effectiveness is presented in Fig. 5A.

First, we note that any  $D$ -optimal design  $\hat{\epsilon}$  must have a diagonal information matrix. In the contrary case, we consider the design

$$\epsilon = \frac{1}{2}\hat{\epsilon} + \frac{1}{2}\tilde{\epsilon},$$

where  $\tilde{\epsilon}$  is a mirror image of the given design  $\hat{\epsilon}$  with respect to zero. The design  $\epsilon$ , as is not difficult to show, has exactly the same determinant for its information matrix as  $\hat{\epsilon}$ . By Theorem 2.2.2 we obtain a design  $\epsilon$ , better than the initial design which contradicts the assumptions about the optimality of the design  $\hat{\epsilon}$ .

For a diagonal dispersion matrix the variance of  $\hat{\eta}(x)$  has the form

$$d(x, \epsilon) = M_{11}^{-1} + M_{22}^{-1}x^2$$

It is easy to obtain that for the given efficiency function, an optimal design<sup>1</sup> is

$$\hat{\epsilon} = \left\{ \begin{array}{llll} x_1 = -1 & x_2 = -\frac{1}{2} & x_3 = \frac{1}{2}, & x_4 = 1 \\ p_1 = \frac{1}{4} & p_2 = \frac{1}{2}, & p_3 = \frac{1}{2}, & p_4 = \frac{1}{4} \end{array} \right\}$$

The graph of  $\lambda(x) d(x, \hat{\epsilon})$  is presented in Fig. 5B.

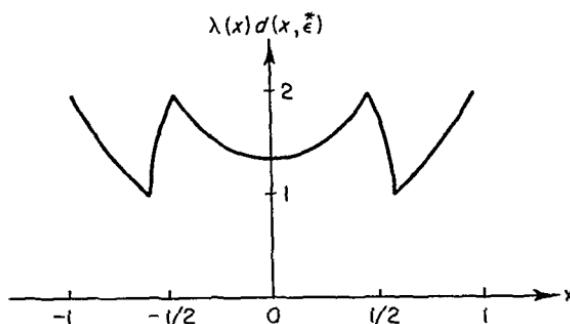
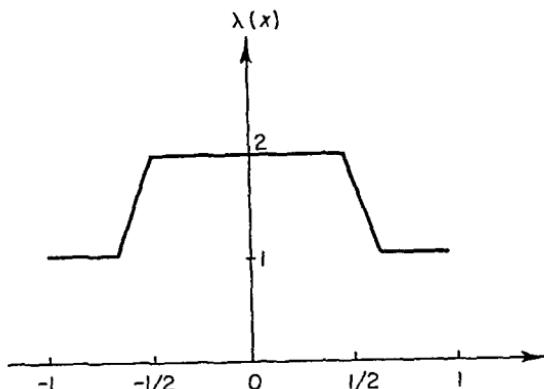
Theorems 2.3.1 and 2.3.2 permit, for many practical important cases, analytic solutions to problems of  $D$ -optimal designs of one-dimensional polynomial regressions [21, 22]. We will assume that  $x$  can belong to one of the three regions  $[-1, 1]$ ,  $[0, \infty]$ , and  $(-\infty, \infty)$ . All of the remaining cases of closed regions can be reduced to one of these three given cases by means of a corresponding translation of the coordinate origin and a change of scale.

**Theorem 2.3.3** Let  $f_\alpha(x) = x^{\alpha+1}$  ( $\alpha = 1, 2, \dots, m$ ), and let  $\lambda(x)$  be one of the following efficiency functions

- (1)  $\lambda(x) \equiv 1, \quad -1 \leq x \leq 1,$
- (2)  $\lambda(x) = (1-x)^{\alpha+1}(1+x)^{\beta+1}, \quad -1 \leq x \leq 1, \alpha > -1, \beta > -1,$
- (3)  $\lambda(x) = \exp(-x), \quad 0 \leq x \leq \infty,$
- (4)  $\lambda(x) = x^{\alpha+1} \exp(-x), \quad 0 \leq x \leq \infty, \alpha > -1,$
- (5)  $\lambda(x) = \exp(-x^2), \quad -\infty \leq x \leq \infty$

Then for each of the cases (1)–(5) the  $D$ -optimal design is unique and

<sup>1</sup> The optimal design is not unique — TRANS



**Fig. 5A.** Step function of efficiency.

**Fig. 5B.** Dependence of  $\lambda(x)d(x_i, \epsilon)$  on  $x$ . At points of the design  $\lambda(x_i)d(x_i, \epsilon) = 2$ ,  $i = 1, 2, 3, 4$ .

is concentrated at  $m$  points with equal weights  $p_i = m^{-1}$ . The points are roots of the polynomials:

- (1)  $(1 - x^2)P'_{m-1}(x)$ , where  $P_m(x)$  is the  $m$ th Legendre polynomial;
- (2)  $P_m^{(\alpha, \beta)}(x)$ , where  $P_m^{(\alpha, \beta)}(x)$  is the  $m$ th Jacobi polynomial with parameters  $\alpha, \beta$ ;
- (3)  $xL_{m-1}^{(1)}(x)$ ;
- (4)  $L_m^{(\alpha)}(x)$ , where  $L_m^{(\alpha)}(x)$  is the  $m$ th Laguerre polynomial with parameter  $\alpha$ ;
- (5)  $H_m(x)$ , where  $H_m(x)$  is the Hermite polynomial.

*Proof.* In all five case the proof is completely analogous. We will consider Case 2. In this case we will follow, in particular, [21], [22], or [21, p. 139].

It is not difficult to see that all points of the  $D$ -optimal design will be interior with respect to the interval  $[-1, 1]$ . Indeed, the quantity

$$\lambda(\tau) d(\tau, \epsilon) = (1 - \tau)^{\alpha+1} (1 + \tau)^{\beta+1} d(\tau, \epsilon)$$

is equal to zero for  $|\tau| = 1$  [ $d(\tau, \epsilon) < \infty$  for finite  $\tau$ ]. It follows that we may consider that the points of the sought design belong to the interval  $[a, b]$ , where

$$a = -1 + \gamma, \quad b = 1 - \gamma, \quad (2.3.11)$$

where  $\gamma$  is some positive number. But in any interval of the type (2.3.11) the system of  $2m$  functions

$$1, (1 - \tau)^{\alpha+1} (1 + \tau)^{\beta+1}, (1 - \tau)^{\alpha+1} (1 + \tau)^{\beta+1} \tau, \dots, (1 - \tau)^{\alpha+1} (1 + \tau)^{\beta+1} \tau^{2m-2}$$

is a Chebyshev system<sup>2</sup> and it follows, from Theorem 2.3.2, that any  $D$ -optimal design for the interval (2.3.11) (which evidently coincides with the  $D$ -optimal design for the interval  $[-1, 1]$ ) consists of  $m$  points. In this case (cf. Corollary 1 of Theorem 2.3.1) the determinant of the information matrix is equal to

$$|M(\epsilon)| = \prod_{i=1}^m p_i \prod_{i=1}^m (1 - x_i)^{\alpha+1} (1 + x_i)^{\beta+1} \prod_{\substack{k < i \\ k=1, 2 \\ k < i}} \prod_{j=i+1}^m (x_k - x_j)^2, \quad (2.3.12)$$

and the optimal allocation is  $p_i = m^{-1}$  ( $i = 1, 2, \dots, m$ ).

In order that the maximum of the quantity

$$T = \prod_{i=1}^m (1 - x_i)^{\alpha+1} (1 + x_i)^{\beta+1} \prod_{\substack{k < i \\ k=1, 2 \\ k < i}} \prod_{j=i+1}^m (x_k - x_j)^2 \quad (2.3.13)$$

hold the condition  $\partial T / \partial x_k = 0$  must be satisfied, or

$$2 \left\{ \frac{1}{x_k - x_1} + \dots + \frac{1}{x_k - x_{k-1}} + \frac{1}{x_k - x_{k+1}} + \dots + \frac{1}{x_k - x_m} \right\} \\ + \frac{\alpha + 1}{x_k - 1} + \frac{\beta + 1}{x_k + 1} = 0 \quad (2.3.14)$$

We set  $f(x) = (\tau - x_1)(\tau - x_2) \dots (\tau - x_m)$ , then (2.3.14) takes on form

$$[f(x_k) f'(x_k)] + [(\alpha + 1) (x_k - 1)] + [(\beta + 1) (x_k + 1)] = 0$$

<sup>2</sup> This statement is incorrect. The proof can be suitably modified. — Trans.

or

$$(1 - x_k^2) f(x) + [\beta - \alpha - (\alpha + \beta + 2)x_k] f'(x_k) = 0.$$

The last equality says that the polynomial

$$(1 - x^2)f(x) + [\beta - \alpha - (\alpha + \beta + 2)x]f'(x)$$

has zeros at the zeros of the polynomial  $f(x)$  and has the same degree. It follows that

$$(1 - x^2)f(x) + [\beta - \alpha - (\alpha + \beta + 2)x]f'(x) = \text{const} \cdot f(x).$$

Comparing the coefficients of  $x^m$ , we find that the constant multiplier equals  $-m(m + \alpha + \beta + 1)$  and the function  $f(x)$  satisfies the differential equation

$$(1 - x^2)f(x) + [\beta - \alpha - (\alpha + \beta + 2)x]f'(x) + m(m + \alpha + \beta + 1)f''(x) = 0. \quad (2.3.15)$$

Equation (2.3.15) has as a solution the Jacobi polynomials  $P_m^{(\alpha, \beta)}(x)$ . It is also well known [24] that (2.3.15) has no other solution, linearly independent of  $P_m^{(\alpha, \beta)}(x)$ , represented in the form of a polynomial. In this way, the function satisfying (2.3.15) with accuracy up to constant multipliers is unique. It follows that it is unique and the  $D$ -optimal design is concentrated on its roots. The theorem is proved.

It is interesting to point out that the problem of maximizing the functional (2.3.12) was solved comparatively long ago in mathematical physics while investigating some spaces of electrostatic systems and is closely tied to the theory of orthogonal polynomials.

For large  $m$  it is possible to talk about the density of the roots of the polynomials. For the polynomials  $(1 - x^2)P_{m-1}(x)$  and  $P_m^{(\alpha, \beta)}$  the following assertion is valid [24].

Let  $[a, b]$  be part of the interval  $[-1, 1]$  and let  $a = \cos \gamma$ ,  $b = \cos \delta$ ,  $\pi \geq \gamma > \delta \geq 0$ . If  $N = N(m; a, b)$  indicates the number of zeros of the polynomial  $(1 - x^2)P_{m-1}(x)$  or  $P_m^{(\alpha, \beta)}(x)$ , lying on the interval  $[a, b]$ , then

$$\lim_m \frac{N(m; a, b)}{m} = |\gamma - \delta|, \quad (2.3.16)$$

or, in other words, the density of the roots  $\rho(x)$  is proportional to  $(1 - x^2)^{-1/2}$

The result (2.3.16) is particularly obvious in its graphical presentation (cf. Fig. 6)

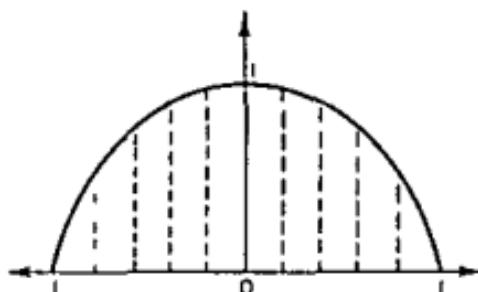
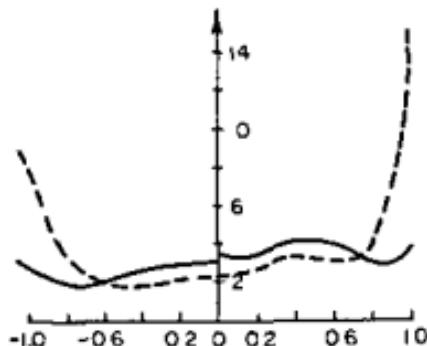


Fig. 6 Position of the roots of orthogonal polynomials for large  $m$

In this way, for a large-degree polynomial regression "the density" of the points of the  $D$ -optimal designs for the interval  $[-1, 1]$  follows a law which is distinct from the uniformly distributed one. The reader can obtain more detailed information about the roots of the polynomial under consideration in [24, 26].

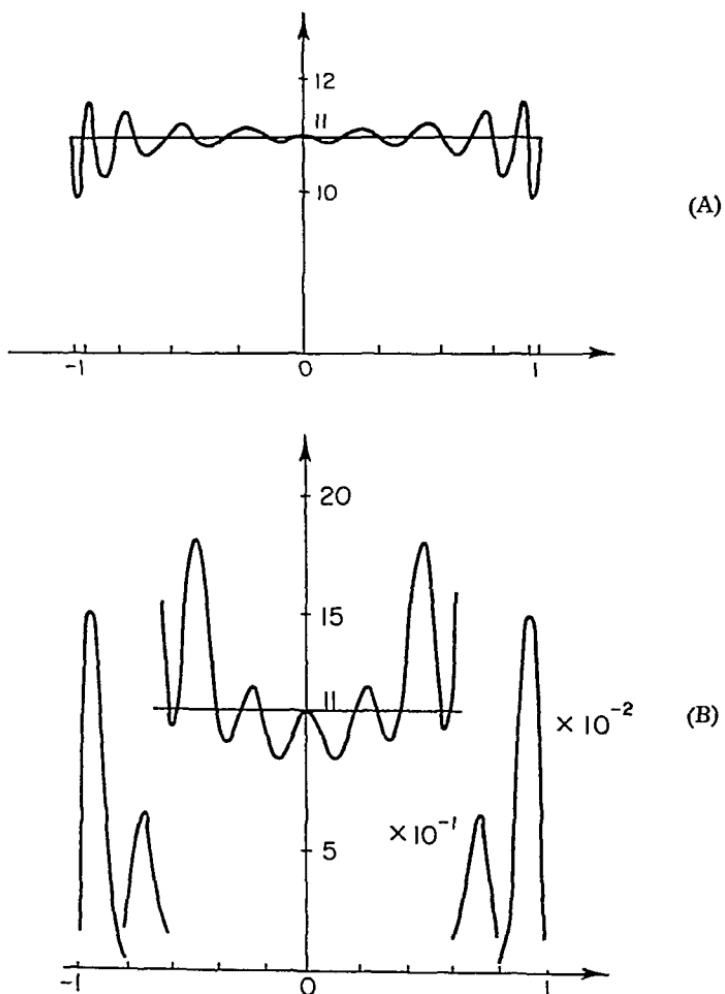
**EXAMPLE** We compare the characteristics of the  $D$ -optimal design  $\hat{\epsilon}$  and the design  $\epsilon$  with the uniform spectrum for polynomials of the second and third degree with  $\lambda(x) = 1$ . Easy calculations show that the ratio of the determinants of the information matrices  $|M(\hat{\epsilon})|$  and  $|M(\epsilon)|$  is equal to  $\sim 3$  for polynomials of the second degree and  $\sim 2.8$  for polynomials of the third degree. In the uniform design (the same in both cases) the number of points was chosen to be 11. The curves  $d(x, \hat{\epsilon})$  and  $d(x, \epsilon)$  are presented in Fig. 7.

Fig. 7 The continuous line corresponds to the variance of the estimate of the curve for the  $D$ -optimal design; the dotted line corresponds to the design with uniform spectrum. The left side of the figure corresponds to quadratic regression, the right to cubic.



EXAMPLE. In those cases when the regression curves  $\eta(x, \theta)$  are polynomials of high order, one can use formula (2.3.16) for an approximate determination of the spectrum of the optimal design. In this case the points of the spectrum can be found by the method depicted in Fig. 6.

In Fig. 8 the dependence of the dispersion  $d(x, \epsilon)$  of the curve



**Fig. 8A.** Variance of the curve  $\hat{\eta}(x, \theta)$  for the approximate  $D$ -optimal design [ $\eta(v, \theta)$  is a tenth-degree polynomial]. The vertical dashes denote the points of the spectrum of the design.

**Fig. 8B.** The same as in Fig. 8A for the uniform design.

$\eta(x, \theta)$  is presented, where  $\eta(x, \theta)$  is a polynomial of the tenth degree ( $m = 11$ ) for the uniform design, consisting of 11 points and of the approximate  $D$ -optimal design  $\epsilon_n$ . We note that  $\max_x d(x, \epsilon_n)$  only insignificantly exceeds  $\max_x d(x, \epsilon) = 11$

## 2.4. Trigonometric Regression on an Interval

I. We consider a regression function appearing as a trigonometric sum of order  $k$

$$\eta(x, \theta) = \theta_0 + \sum_{n=1}^k [\theta_n \cos nx + \omega_n \sin nx] \quad (2.4.1)$$

The domain of the measurement  $x$  is chosen equal to  $[0, 2\pi]$ . We will assume that the function of effectiveness  $\lambda(x)$  is constant in this domain. Above it was shown that all  $D$ -optimal designs have one and only one information matrix  $\hat{M}$ . If some  $D$ -optimal design is found (this is possible with a very large number of points), then the remaining optimal designs (if any exist) can be found by solving the system of equations

$$\sum_{i=1}^n p_i f_\alpha(x_i) f_\beta(x_i) = \hat{M}_{\alpha\beta} \quad (1 \leq \alpha, \beta \leq k) \quad (2.4.2)$$

$$\lambda(x_i) f'(x_i) f'(x_i) = m - 2k + 1 \quad (i = 1, 2, \dots, n)$$

It is not difficult to show that the system (2.4.2) is sufficient to define the coordinates  $x_i$  of the design and the weights  $p_i$  for  $n \leq m(m+1)/2$ .

Despite its semi-intuitive character, such an approach is sometimes adequate for constructing  $D$ -optimal designs. Indeed, this approach will be used in the present section.

### II. Since the system of functions

$$f(x) = [1 \ sin x \ cos x \ \dots \ sin kx \ cos kx] \quad (2.4.3)$$

is orthogonal on the interval  $[0, 2\pi]$ , the design  $\epsilon_c$  with allocation density  $p(x) = (2\pi)^{-1}$  will have a diagonal information matrix,

$$\begin{aligned} M_{11} &= 1, \\ M_{\alpha\alpha} &= (2\pi)^{-1} \int_0^{2\pi} f_\alpha(x) f_\alpha(x) dx = \frac{1}{2}, \quad \alpha > 1, \\ M_{\alpha\beta} &= (2\pi)^{-1} \int_0^{2\pi} f_\alpha(x) f_\beta(x) dx = 0, \quad \alpha \neq \beta \end{aligned} \quad (2.4.4)$$

Recall that

$$\int_0^{2\pi} \sin kx \cos mx dx = 0; \quad \int_0^{2\pi} \sin^2 kx dx = \pi; \quad \int_0^{2\pi} \cos^2 kx dx = \pi.$$

We compute the dispersion  $d(x, \epsilon_c)$ :

$$D_{11} = M_{11}^{-1} = 1, \quad D_{\alpha\alpha} = M_{\alpha\alpha}^{-1} = 2, \quad \alpha > 1,$$

$$d(x, \epsilon_c) = 1 + 2 \sum_{\alpha=1}^m f_{\alpha}^2(x) = 1 + 2 \sin^2 x + 2 \cos^2 x + \cdots + 2 \sin^2 kx + 2 \cos^2 kx = 2k + 1 = m. \quad (2.4.5)$$

From (2.4.5) and the equivalence theorem it follows that the uniform design is  $D$ -optimal in the case of trigonometric regression with  $\lambda(x) = 1$ . It is a remarkable fact that the dispersion of the regression curve is the same for any point belonging to the interval  $[0, 2\pi]$ .

We now take up the problem of finding the design (or designs) with a small number of points at which it is necessary to take the measurements. According to (2.4.2) the coordinates of the points of the design  $x_i$  and weights  $p_i$  ( $i = 1, 2, \dots, n$ ) must be such that the following equations hold:

$$\sum_{i=1}^n p_i = 1,$$

$$\sum_{i=1}^n p_i f_{\alpha}^2(x_i) = \frac{1}{2}, \quad \alpha > 1, \quad (2.4.6)$$

$$\sum_{i=1}^n p_i f_{\alpha}(x_i) f_{\beta}(x_i) = 0, \quad \alpha \neq \beta.$$

In (2.4.6) the last  $n$  equations of the system (2.4.2) are not written out. In the given case one cannot extract any information about the  $D$ -optimal design from these equations, since they are satisfied identically for any  $x$ .

It is not difficult to verify that for equal weights  $p_i = n^{-1}$  and

$$x_i = 2\pi[(i - 1)/n], \quad i = 1, 2, \dots, n, \quad n \geq 2k + 1, \quad (2.4.7)$$

equations (2.4.6) are satisfied. It follows that the corresponding design is  $D$ -optimal.

The assumed distribution of the points of the design is by far not the unique optimal placement of points. Thus, for example, for equal weights and

$$x_i = 2\pi[(i-1)/n] + \varphi, \quad i = 1, 2, \dots, n, \quad n \geq 2k+1, \quad (2.4.8)$$

the designs are also  $D$ -optimal. In this manner, for trigonometric regression, any design with uniform spectrum, containing no less than  $2k+1$  points and equal weight at each point, is  $D$ -optimal. The case  $n = 2k+2$  was first considered in [27].

### III. We consider the regression function

$$\eta(x, \theta) = \theta_0 + \sum_{a=1}^{k-1} [\theta_a \cos ax + \omega_a \sin ax] \quad (2.4.9)$$

Let the design

$$\dot{\epsilon}_k = \{x_i, p_i (i = 1, 2, \dots, n)\}$$

be  $D$ -optimal for the trigonometric regression (2.4.1) of order  $k$ . Then from (2.4.6) it follows that

$$\sum_{i=1}^n p_i = 1$$

$$\sum_{i=1}^n p_i f_\alpha^2(x_i) = \frac{1}{2} \quad \alpha > 1, \quad (2.4.10)$$

$$\sum_{i=1}^n p_i f_\alpha(x_i) f_\beta(x_i) = 0 \quad \alpha \neq \beta,$$

$$\alpha, \beta \leq m - 2 = 2k - 1$$

From (2.4.10) and the definition of the information matrix we find that for the trigonometric regression (2.4.9) of order  $k-1$ ,

$$D_{11} = M_{11}^{-1} = 1, \quad D_{\alpha\alpha} = M_{\alpha\alpha}^{-1} = 2, \quad 1 \leq \alpha \leq 2k-1$$

From this, we find that

$$\begin{aligned} d(x | \dot{\epsilon}_k) &= 1 + 2 \cos^2 x + 2 \sin^2 x + \\ &+ 2 \cos^2(k-1)x + 2 \sin^2(k-1)x = 2k-1 \end{aligned}$$

From this it follows that the design which is *D*-optimal for trigonometric regression of the  $k$  order is also *D*-optimal for trigonometric regression of order  $k - 1$ , and for any trigonometric regression of order less than  $k$ .

## 2.5. Computational Methods for Construction of *D*-Optimal Designs

I. An analytic solution of the problem of constructing *D*-optimal designs is possible only in the simplest cases (cf. Sections 2.2–2.4) and requires a special approach in each distinct case. Numerous optimal functions are presented in the literature of designing *D*-optimal experiments. A table of examples of *D*-optimal designs appears in preprint number 16.<sup>3</sup> In practically all of the works in the literature, the optimal design is constructed using semi-intuitive considerations (e.g., from the consideration of symmetry of the distribution of the points of the design, orthogonality of the functions  $f_\alpha(x)$  over the points of the design, etc; see e.g., Section 2.4). The designs obtained in this manner are verified afterward by the equivalence theorem of *D*-optimality. Such methods are fruitful when the domain  $X$  of possible values of the control variables  $x$  is a simple geometric figure (e.g., a sphere or a cube in the space of control variables) and the efficiency function  $\lambda(x)$  is constant on  $X$ . It is natural that in the more complicated situations such methods are hardly useful. Just as in many other mathematical problems, complication of the analytical form of the functionals under investigation [here  $\eta(x, \theta)$ ,  $\lambda(x)$ ,  $|M(\epsilon)|$ ,  $d(x, \epsilon)$ ] makes it necessary to turn to numerical methods for finding the solution. From the conceptual point of view one of the simplest methods would be a direct search for the maximum of the determinant  $|M(\epsilon)|$ . The mathematical apparatus for a numerical search of the conditional extremum of some function

$$\mathcal{L}(x_1, x_2, \dots, x_n, p_1, p_2, \dots, p_n) = |M(\epsilon)|$$

$$\sum_{i=1}^n p_i = 1 \quad (2.5.1)$$

is sufficiently well developed at the present time and is broadly applied

<sup>3</sup> This is presumably number 16 of the preprint series of the Laboratory of Statistical Research of Moscow State University.—TRANS.

for the solution of many problems (cf., e.g., [17, 28, 29]). There are many programs for electronic computing machines using various computational methods (e.g., the method of random search, all possible modifications of gradient methods, methods of descent, etc.). However, for the solution of the extremal problem (2.5.1) these methods exhaust themselves rather rapidly, since the collection of variables with respect to which it is necessary to seek the extremum grows very quickly with the growth of the number of estimated parameters  $m$ . It is easy to determine (cf. Theorem 2.2.3) that the dimension of the space of variables, for the extremal problem for finding  $D$ -optimal designs, can be contained between the limits

$$km + m \leq N \leq k[m(m+1)/2] + [m(m+1)/2]$$

where  $k$  is the dimension of the space of control variables  $x$ . Practice shows that for the average electronic computing machine (e.g., M-20) an upper bound  $N$  is somewhere in the limits 10–30 (to a significant degree, this boundary depends on the form of the functional for which the extremum is being sought).

Such a bound on the existing methods requires us to turn to the construction of special numerical methods for obtaining  $D$ -optimal designs. It is clear that constructional methods, significantly exceeding their usual performance, necessitates relying on concrete properties of the functionals under investigation [30].

**II** The basic idea used for constructing such methods consists in the following. As was shown in Section 2.2,

$$(d/d\alpha) \log |M(\epsilon_1)| = \text{Tr}\{(1-\alpha)M(\epsilon_0) + \alpha M[\epsilon(x)]\}^{-1}(M[\epsilon(x)] - M(\epsilon_0)) \quad (2.5.2)$$

where  $\epsilon_1 = (1-\alpha)\epsilon_0 + \alpha\epsilon(x)$ ,  $M(\epsilon_0)$  is the information matrix of some nondegenerate design  $\epsilon_0$ , and  $M[\epsilon(x)]$  is the information matrix of the design  $\epsilon(x)$ , consisting of one point  $x$ . For  $\alpha = 0$ ,

$$(d/d\alpha) \log |M(\epsilon_1)||_{\alpha=0} = \lambda(x) d(x, \epsilon_0) - m,$$

or, for sufficiently small  $\alpha$ ,

$$\log |M(\epsilon_1)| \approx \log |M(\epsilon_0)| + u[\lambda(x) d(x, \epsilon_0) - m] \quad (2.5.3)$$

By Theorem 2.2.1

$$\max_x \lambda(x) d(x, \epsilon_0) - m = \delta \geq 0, \quad (2.5.4)$$

and equality holds only if the design  $\epsilon_0$  is  $D$ -optimal.

Let the design  $\epsilon_0$  not be  $D$ -optimal. Mixing part of the allocation at the point  $x_0$ , corresponding to  $\max_x \lambda(x) d(x, \epsilon_0)$ , we obtain, for small  $\alpha$ ,

$$\log |M(\epsilon_1)| \simeq \log |M(\epsilon_0)| + \alpha_0 \delta > \log |M(\epsilon_0)|. \quad (2.5.5)$$

Inequality (2.5.5) says that the design  $\epsilon_1 = (1 - \alpha_0)\epsilon_0 + \alpha_0 \epsilon(x)$  with information matrix  $M(\epsilon_1)$  is better than the design  $\epsilon_0$ , i.e., the determinant  $|M(\epsilon_1)| \geq |M(\epsilon_0)|$ .

Constructing the design  $\epsilon_1$ , we find the point  $x_1$ , corresponding to  $\max_x \lambda(x) d(x, \epsilon_1)$ . Mixing a small part of the allocation  $\alpha_1$  at this point we construct a design better than the design  $\epsilon_1$ . Continuing this procedure we obtain a sequence of matrices  $M(\epsilon_0), M(\epsilon_1), \dots, M(\epsilon_s)$  such that

$$|M(\epsilon_0)| < |M(\epsilon_1)| \leq \dots \leq |M(\epsilon_s)|. \quad (2.5.6)$$

The sequence (2.5.6) is bounded above:

$$|M(\epsilon_s)| \leq |M(\hat{\epsilon})|,$$

where  $\hat{\epsilon}$  is the  $D$ -optimal design. Therefore, for a suitable choice  $\alpha_s$  it converges to  $|M(\hat{\epsilon})|$  ( $\alpha_s$  must converge to zero as  $s \rightarrow \infty$ , but not too rapidly, because the sequence  $\{|M(\epsilon_s)|\}$  might converge without having attained its upper bound).

**III.** We will reduce the rough outline of the numerical method of constructing  $D$ -optimal designs, presented in the preceding section, to concrete iterative procedures, and we will prove their convergence.

First, we will generalize formula (2.5.3) to the case of arbitrary  $\alpha$ , contained in the interval  $(0, 1)$ .

**Lemma 2.5.1.** *Let  $M$  be a nondegenerate  $m \times m$  matrix and let  $F$  be an  $m \times k$  matrix; then*

$$|M + FF'| = |M| |I_k + F'M^{-1}F|. \quad (2.5.7)$$

*Proof* It is known (cf. Theorem 1.1.3) that

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| |D - CA^{-1}B| = |A - BD^{-1}C| |D|$$

Setting  $A = M$ ,  $B = F$ ,  $C = -F$ , and  $D = I_k$ , we obtain the necessary result

$$|M + FF'| = |M| |I_k + FM^{-1}F|$$

**Theorem 2.5.1.** Let  $M(\epsilon_s)$  be the information matrix of a nondegenerate design  $\epsilon_s$  and  $M[\epsilon(x)]$  be the information matrix of the design concentrated at the single point  $x$ . Then the determinant of the information matrix of a linear combination of these designs  $\epsilon_{s+1} = (1 - \alpha)\epsilon_s + \alpha\epsilon(x)$ , equals

$$|M(\epsilon_{s+1})| = (1 - \alpha)^m \{1 + [\alpha(1 - \alpha)] \lambda(x) d(x, \epsilon_s)\} |M(\epsilon_s)| \quad (2.5.8)$$

*Proof* By definition,

$$\begin{aligned} M(\epsilon_{s+1}) &= (1 - \alpha) M(\epsilon_s) + \alpha \lambda(x) f(x) f'(x) \\ &= (1 - \alpha) \{M(\epsilon_s) + [\alpha(1 - \alpha)] \lambda(x) f(x) f'(x)\} \end{aligned}$$

Setting  $F = [\alpha\lambda(x)/(1 - \alpha)]^{1/2} f(x)$  in (2.5.7) we obtain

$$|M(\epsilon_{s+1})| = (1 - \alpha)^m \{1 + [\alpha\lambda(x)/(1 - \alpha)] f(x) M^{-1}(\epsilon_s) f(x)\} |M(\epsilon_s)|$$

Since  $f(x) M^{-1}(\epsilon_s) f(x) = d(x, \epsilon_s)$ , it follows that

$$|M(\epsilon_{s+1})| = (1 - \alpha)^m \{1 + [\alpha/(1 - \alpha)] \lambda(x) d(x, \epsilon_s)\} |M(\epsilon_s)|$$

The theorem is proved

From (2.5.8) it is not difficult to see that the determinant of the information matrix of  $\epsilon_{s+1}$  depends on the quantity  $\alpha$  and on the coordinates of the point at which the design  $\epsilon(x)$  is concentrated. We will show that it is always possible to find a point  $x$  and an  $\alpha$  for which  $|M(\epsilon_{s+1})| > |M(\epsilon_s)|$ , if the design  $\epsilon_s$  is not  $D$  optimal.

**Theorem 2.5.2.** *The largest possible value of  $|M(\epsilon_{s+1})|$  for a given  $\epsilon_s$  equals*

$$\begin{aligned} \max_{x,\alpha} |M(\epsilon_{s+1})| &= \left[ \frac{\lambda(x_s) d(x_s, \epsilon_s)}{m} \right]^m \\ &\times \left[ \frac{m-1}{\lambda(x_s) d(x_s, \epsilon_s) - 1} \right]^{m-1} |M(\epsilon_s)| > |M(\epsilon_s)|, \end{aligned} \quad (2.5.9)$$

where  $x_s$  is a solution to the equation  $\lambda(x_s) d(x_s, \epsilon_s) = \max_x \lambda(x) d(x, \epsilon_s)$ .

*Proof.* From (2.5.8) it is easy to see that for any  $\alpha$ , the determinant  $|M(\epsilon_{s+1})|$  is an increasing function of  $\lambda(x) d(x, \epsilon_s)$ . From this it obviously follows that the design  $\epsilon(x)$  must be concentrated at the point  $x$ , corresponding to  $\max_x \lambda(x) d(x, \epsilon_s)$ . Therefore  $\max_{x,\alpha} |M(\epsilon_{s+1})|$  will hold when  $x = x_s$  and  $\alpha$  satisfies the equation

$$\frac{\partial}{\partial \alpha} \log |M(\epsilon_{s+1})| = \frac{\lambda(x_s) d(x_s, \epsilon_s) - 1}{1 - \alpha + \alpha \lambda(x_s) d(x_s, \epsilon_s)} - \frac{m-1}{1 - \alpha} = 0. \quad (2.5.10)$$

The solution of (2.5.10) is

$$\alpha_s = \frac{\lambda(x_s) d(x_s, \epsilon_s) - m}{[\lambda(x_s) d(x_s, \epsilon_s) - 1]m}. \quad (2.5.11)$$

Since, by assumption, the design  $\epsilon_s$  is not D-optimal, by Theorem 2.2.1

$$\lambda(x_s) d(x_s, \epsilon_s) - m = \max_x [\lambda(x) d(x, \epsilon_s) - m] > 0,$$

and it follows that  $\alpha_s > 0$ . On the other hand, by means of a simple differentiation it is easy to verify that

$$\left( \frac{\partial^2}{\partial \alpha^2} \log |M(\epsilon_{s+1})| \right) \Bigg|_{\substack{\alpha=\alpha_s \\ x=x_s}} < 0.$$

Therefore  $\alpha_s$  corresponds to the maximum of  $\log |M(\epsilon_{s+1})|$ . Therefore

$$\max_{\alpha,x} \log |M(\epsilon_{s+1})| > \log |M(\epsilon_s)|. \quad (2.5.12)$$

Setting  $\alpha_s$ , defined by formula (2.5.11), into (2.5.8) we obtain, after a simple calculation

$$|M(\epsilon_{s+1})| \Bigg|_{\substack{\alpha=\alpha_s \\ x=x_s}} = \left[ \frac{\lambda(x_s) d(x_s, \epsilon_s)}{m} \right]^m \left[ \frac{m-1}{\lambda(x_s) d(x_s, \epsilon_s) - 1} \right]^{m-1} |M(\epsilon_s)|. \quad (2.5.13)$$

Combining (2.5.12) and (2.5.13) we obtain the validity of the theorem

We consider the following iterative procedure. Given some non-degenerate design  $\epsilon_0$  (initial approximation)

$$\epsilon_0 = \left\{ \begin{array}{l} x_1, x_2, \dots, x_n \\ p_1, p_2, \dots, p_n \\ \sum_{i=1}^n p_i = 1, \quad n \geq m \end{array} \right\} \quad (2.5.14)$$

1 We compute its information matrix

$$M(\epsilon_0) = \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i)$$

and its inverse, the dispersion matrix  $D(\epsilon_0)$

2 The point  $x_0$  is found at which  $\max_x \lambda(x) d(x, \epsilon_0)$  is attained, and the corresponding value  $\lambda(x_0) d(x_0, \epsilon_0)$  is determined

3 The design  $\epsilon_1 = (1 - \alpha_0) \epsilon_0 + \alpha_0 \epsilon(x_0)$  is constructed, or in more detail

$$\epsilon_1 = \left\{ \frac{x_1}{(1 - \alpha_0)p_1}, \frac{x_2}{(1 - \alpha_0)p_2}, \dots, \frac{x_n}{(1 - \alpha_0)p_n}, \frac{x_0}{\alpha_0} \right\} \quad (2.5.15)$$

"The step"  $\alpha_0$  is chosen such that the increase in the determinant of the information matrix is maximal, that is (cf. Theorem 2.5.2),

$$\alpha_0 = \delta_0 [\delta_0 + (m - 1)]^{-1} \quad (2.5.16)$$

where  $\delta_0 = \lambda(x_0) d(x_0, \epsilon_0) - m$

4 The information matrix  $M(\epsilon_1)$  of the design  $\epsilon_1$  is computed along with its inverse matrix  $D(\epsilon_1)$

After the matrix  $D(\epsilon_1)$  is found, operations 2-4 are repeated with  $\epsilon_0$  replaced by  $\epsilon_1$ , these operations are then repeated with  $\epsilon_2, \epsilon_3, \text{ etc}$

We show that the iterative procedure outlined converges and that its limit design  $\epsilon$  coincides with one of the  $D$ -optimal designs

**Theorem 2.5.3** *Let the conditions of the equivalence theorem be satisfied, then the iterative procedure 1-4 converges, in which case*

$$\lim_{s \rightarrow \infty} |M(\epsilon_s)| = |M(\epsilon)|$$

where  $M(\hat{\epsilon})$  is the information matrix corresponding to the  $D$ -optimal design.

*Proof.* Let the design  $\epsilon_0$  not be  $D$ -optimal. Then, in view of Theorem 2.5.2 and the determinant of the  $D$ -optimal design,

$$|M(\epsilon_0)| < |M(\epsilon_1)| \leq \dots \leq |M(\epsilon_s)| \leq |M(\hat{\epsilon})|. \quad (2.5.17)$$

But, as is known, any bounded monotone nondecreasing sequence converges. It follows that the sequence  $|M(\epsilon_0)|, |M(\epsilon_1)|, \dots, |M(\epsilon_s)|$  converges to some limit  $|M(\bar{\epsilon})|$ . Therefore, for the proof of the theorem it is sufficient to show that

$$|M(\bar{\epsilon})| = |M(\hat{\epsilon})|. \quad (2.5.18)$$

We will assume the contrary:

$$|M(\bar{\epsilon})| < |M(\hat{\epsilon})|. \quad (2.5.19)$$

In view of the convergence of the sequence  $|M(\epsilon_0)|, |M(\epsilon_1)|, \dots, |M(\epsilon_s)|$ , for any small positive number  $\gamma$  there is an  $\bar{s}$  such that for any  $s > \bar{s}$  the following inequality holds:

$$|M(\epsilon_{s+1})| - |M(\epsilon_s)| \leq \gamma,$$

or by Theorem 2.5.2,

$$|M(\epsilon_s)| \left[ \left( \frac{\delta_s + m}{m} \right)^m \left( \frac{m-1}{\delta_s + (m-1)} \right)^{m-1} - 1 \right] \leq \gamma. \quad (2.5.20)$$

Inequality (2.5.20) can be rewritten in the form

$$\Psi(\delta_s) = \left( \frac{\delta_s + m}{m} \right)^m \left( \frac{m-1}{\delta_s + (m-1)} \right)^{m-1} < 1 + \gamma_1,$$

where  $\gamma_1 = |M(\epsilon_s)|^{-1}\gamma$ . The function  $\Psi(\delta_s)$  is increasing for  $\delta_s > 0$ ; it follows that for any  $\Delta > 0$  there is a  $\gamma$  such that

$$\lambda(x_s) d(x_s, \epsilon_s) - m = \delta_s \leq \Delta.$$

In this way we can always specify  $\bar{s}(\Delta)$  such that  $\lambda(x_s) d(x_s, \epsilon_s) - m$  will be less than any given positive number  $\Delta$ . But by assumption (2.5.19) and the equivalence theorem 2.2.1, for any  $s$ ,

$$\lambda(x_s) d(x_s, \epsilon_s) - m \geq \xi > 0.$$

Choosing  $\Delta < \xi$ , we arrive at the contradiction, which proves our theorem

**IV.** The iterative process considered can terminate if one of the following conditions is satisfied

$$\begin{aligned} \alpha_s &< \gamma_1, \\ \frac{|M(\epsilon_{s+1})| - |M(\epsilon_s)|}{|M(\epsilon_{s+1})|} &< \gamma_2, \\ m^{-1}[\lambda(x_s) \lambda(x_s, \epsilon_s) - m] &< \gamma_3 \end{aligned} \quad (2.5.21)$$

Since the information matrix is identical for all  $D$ -optimal designs for the given regression problem, the left-hand sides of inequalities (2.5.21) are intimately related to one another. All three rules of terminating the computation are equivalent in practice and the choice of any of these stipulates a concrete program for carrying out the computation.

**EXAMPLE** In Fig. 9A values of  $|D(\epsilon_s)|$  are plotted corresponding to the first several iterations for constructing the optimal design for the regression problem

$$\begin{aligned} \eta(x, \theta) &= \theta_1 + \theta_2 x + \theta_3 x^2, \\ \lambda(x) &= (1-x)^2, \quad -1 \leq x \leq 1 \end{aligned}$$

In Fig. 9B the graph of the function  $\lambda(x) d(x, \epsilon_s)$  is plotted for several choices of these iterations.

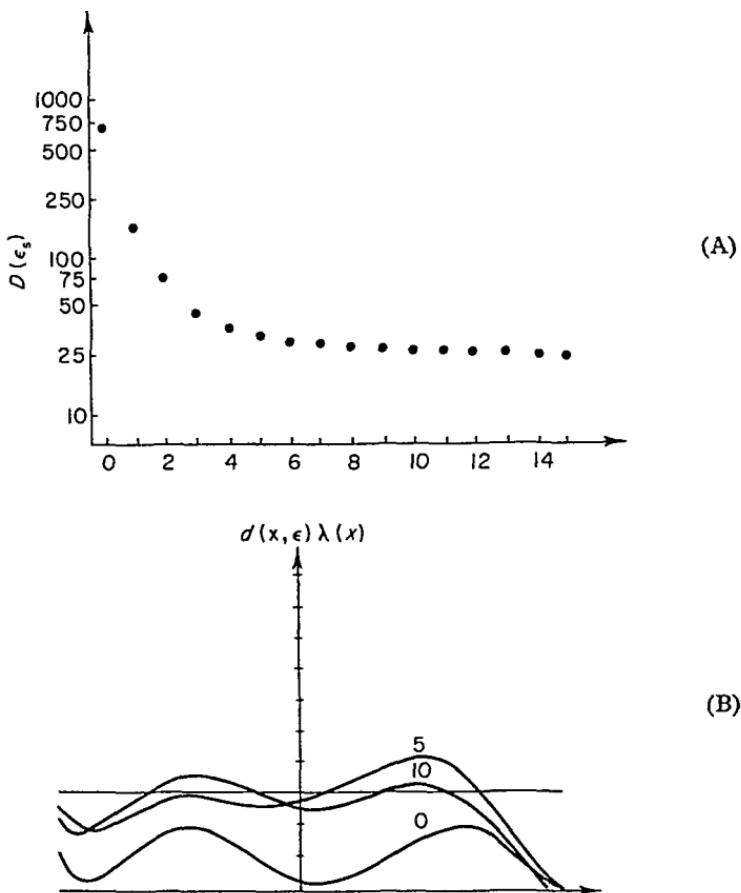
As an initial design the following design was chosen

$$\epsilon_0 = \left\{ \begin{array}{lll} x_1 = -\frac{1}{2} & x_2 = 0 & x_3 = \frac{1}{2} \\ p_1 = \frac{1}{3} & p_2 = \frac{1}{3} & p_3 = \frac{1}{3} \end{array} \right\}$$

The maximal value of the functional  $\lambda(x) d(x, \epsilon_s)$  is found by a lattice method with step  $\Delta x = 0.02$ .

## 2.6. Some Particular Iterative Procedures for Constructing $D$ -Optimal Designs

**I** The most cumbersome operations in the iterative procedure presented are the inversion of the information matrix (its dimension



**Fig. 9A.** Values of the determinant  $|D(\epsilon_s)|$ .

**Fig. 9B.** Variation of the form of the function  $\lambda(x) d(x, \epsilon_s)$  with increasing  $s$ . The number near each curve indicates the number of iterations.

is  $m \times m$ , where  $m$  is the number of unknown parameters) and finding the maximal value of the product  $\lambda(x) d(x, \epsilon_s)$  in a  $k$ -dimensional closed region  $X$ .

It will be shown that the operation of inverting the matrix can be transformed to a significantly simpler operation.

### Lemma 2.6.1

$$(I_p + AB)^{-1} = I_p - A(I_q + BA)^{-1} B, \quad (2.6.1)$$

where  $I_p$  and  $I_q$  are identity matrices of dimension  $p \times p$  and  $q \times q$ ,  $A$  is a  $p \times q$  matrix, and  $B$  is a  $q \times p$  matrix.

*Proof* We multiply both sides of (2.6.1) on the right by the matrix  $I_p + AB$ , and we obtain

$$\begin{aligned} I_p &= I_p(I_s + AB) - A(I_s + BA)^{-1}B(I_s + AB) \\ &= I_p + AB - A(I_s + BA)^{-1}(B + BAB) \\ &= I_p + AB - A(I_s + BA)^{-1}(I_s + BA)B \\ &= I_p + AB - AB = I_p \end{aligned}$$

The lemma is proved

**Theorem 2.6.1.** Let the design  $\epsilon_{s+1}$  be a linear combination of the nondegenerate design  $\epsilon_s$  and the design  $\epsilon(x)$ , concentrated at one point  $x$ :  $\epsilon_{s+1} = (1 - \alpha)\epsilon_s + \alpha\epsilon(x)$ . Then

(1) The dispersion matrix  $D(\epsilon_{s+1})$  of the design  $\epsilon_{s+1}$  is related to the dispersion matrix  $D(\epsilon_s)$  of the design  $\epsilon_s$  by the relationship

$$D(\epsilon_{s+1}) = (1 - \alpha)^{-1} \left[ I - \frac{\alpha\lambda(x) D(\epsilon_s) f(x) f'(x)}{1 - \alpha + \alpha\lambda(x) d(x - \epsilon_s)} \right] D(\epsilon_s) \quad (2.6.2)$$

(2) The dispersion  $d(\hat{x}, \epsilon_{s+1})$  of the estimate of the response surface for the design  $\epsilon_{s+1}$  is expressed through the dispersion  $d(\hat{x}, \epsilon_s)$  of the estimate of the response surface for the design  $\epsilon_s$  in the following way

$$d(\hat{x}, \epsilon_{s+1}) = (1 - \alpha)^{-1} \left[ d(\hat{x}, \epsilon_s) - \frac{\alpha\lambda(x) d^2(x, \hat{x}, \epsilon_s)}{1 - \alpha + \alpha\lambda(x) d(x - \epsilon_s)} \right] \quad (2.6.3)$$

*Proof* By definition

$$M(\epsilon_{s+1}) = (1 - \alpha) M(\epsilon_s) + \alpha M[\epsilon(x)],$$

or

$$\begin{aligned} M^{-1}(\epsilon_{s+1}) &= \{(1 - \alpha) M(\epsilon_s) + \alpha M[\epsilon(x)]\}^{-1} \\ &= (1 - \alpha)^{-1} \{I_m + [\alpha/(1 - \alpha)] M^{-1}(\epsilon_s) M[\epsilon(x)]\}^{-1} M^{-1}(\epsilon_s) \end{aligned}$$

Considering that  $D(\epsilon_s) = M^{-1}(\epsilon_s)$  and  $M[\epsilon(x)] = \lambda(x)f(x)f'(x)$  we rewrite the given expression in the form

$$D(\epsilon_{s+1}) = (1 - \alpha)^{-1} \{I_m + [\alpha\lambda(x)(1 - \alpha)] D(\epsilon_s) f(x) f'(x)\}^{-1} D(\epsilon_s)$$

We set

$$[\alpha\lambda(x)/(1 - \alpha)] D(\epsilon_s) f(x) = A \quad \text{and} \quad f'(x) = B$$

and make use of Lemma 2.6.1:

$$\begin{aligned} & \{I_m + [\alpha\lambda(x)/(1 - \alpha)] D(\epsilon_s) f(x_s) f'(x_s)\}^{-1} \\ &= I_m - [\alpha\lambda(x)/(1 - \alpha)] D(\epsilon_s) f(x) \\ & \quad \times \{1 + f'(x)[\alpha\lambda(x)/(1 - \alpha)] D(\epsilon_s) f(x)\}^{-1} f'(x). \end{aligned} \quad (2.6.4)$$

Since  $f'(x) D(\epsilon_s) f(x) = d(x, \epsilon_s)$ , and correspondingly the expression in brackets in the right-hand side of (2.6.4) is a scalar quantity, then

$$\left[ I_m + \frac{\alpha\lambda(x)}{1 - \alpha} D(\epsilon_s) f(x) f'(x) \right]^{-1} = I_m - \frac{\alpha\lambda(x) D(\epsilon_s) f(x) f'(x)}{1 - \alpha + \alpha\lambda(x) d(x, \epsilon_s)}.$$

From this it follows that

$$D(\epsilon_{s+1}) = (1 - \alpha)^{-1} \left[ I - \frac{\alpha\lambda(x) D(\epsilon_s) f(x) f'(x)}{1 - \alpha + \alpha\lambda(x) d(x, \epsilon_s)} \right] D(\epsilon_s),$$

which is what was required to be shown.

(2) Multiplying (2.6.2) on the left and on the right respectively by  $f'(\tilde{x})$  and  $f(\tilde{x})$ , we obtain

$$\begin{aligned} & f'(\tilde{x}) D(\epsilon_{s+1}) f(\tilde{x}) \\ &= (1 - \alpha)^{-1} \left[ f'(\tilde{x}) D(\epsilon_s) f(\tilde{x}) - \frac{\alpha\lambda(x) f'(\tilde{x}) D(\epsilon_s) f(x) f'(x) D(\epsilon_s) f(\tilde{x})}{1 - \alpha + \alpha\lambda(x) d(x, \epsilon_s)} \right]. \end{aligned} \quad (2.6.5)$$

Making use of the fact that

$$d(x, \tilde{x}, \epsilon) = f'(\tilde{x}) D(\epsilon) f(x) = f'(x) D(\epsilon) f(\tilde{x})$$

[we note that  $d(x, \tilde{x}, \epsilon)$  is the covariance of the estimate  $\hat{\eta}(x, \theta)$  and  $\hat{\eta}(\tilde{x}, \theta)$ ], we rewrite (2.6.5) in the form

$$d(\hat{x}, \epsilon_{s+1}) = (1 - \alpha)^{-1} \left[ d(\hat{x}, \epsilon_s) - \frac{\alpha\lambda(x) d^2(x, \hat{x}, \epsilon_s)}{1 - \alpha + \alpha\lambda(x) d(x, \epsilon_s)} \right].$$

The theorem is proved.

Formulas (2.6.2) and (2.6.3) permit us to avoid inversion of the information matrix  $M(\epsilon_s)$  with the exception of the zeroth iteration. For a large number of unknown parameters this considerably reduces

the volume of computation. At the same time, the accuracy of determining the elements of the dispersion matrix is increased. The increased accuracy of the calculations when computing, for example, on an electronic computing machine is necessitated by the fact that the operations of multiplication, necessary for inversion of matrices, are replaced by operations of summation. In this way it is recommended that the first and second stages of each iteration be carried out with the use of formulas (2.6.2) and (2.6.3), respectively.

The following interesting result follows from Theorem 2.6.1. If  $\alpha_s$  is chosen according to (2.5.16), then the dispersion at the point  $x_s$  for the design  $\epsilon_{s+1} = (1 - \alpha_s)\epsilon_s + \alpha_s\epsilon(x_s)$  remains equal (cf. also Fig. 9B) to the number of unknown parameters

$$\lambda(x_s) d(x_s, \epsilon_{s+1}) = \frac{\lambda(x_s) d(x_s, \epsilon_s)}{1 - \alpha_s + \alpha_s \lambda(x_s) d(x_s, \epsilon_s)} = m$$

This fact says that further transfer of resources to the point  $x_s$  becomes unwise, since obviously a point  $x_{s+1}$  can be found for which  $\lambda(x_{s+1}) d(x_{s+1}, \epsilon_{s+1}) > m$  and it follows that it is more advantageous to transfer at least part of the allocation  $\alpha$  to this point.

II. In the second step of each iteration it is necessary to find  $\max_x \lambda(x) d(x, \epsilon_s)$ . From (2.5.13) it is not difficult to see that the  $(s+1)$ st iteration will be more effective if we succeed in finding the absolute maximum of the function  $\lambda(x) d(x, \epsilon_s)$ . The given function, as a rule, will have (cf. Fig. 9B) no less than  $m$  local maxima. Unfortunately the majority of the standard programs for doing this seek only local maxima of the function under study.

The problem of finding the absolute maxima, particularly for large-dimensional spaces of control variables, necessitates making the existing programs more complicated and leads to significantly increased computational time. Therefore, in the majority of cases the iterative procedure requires less total time (despite the comparatively low efficiency of each iteration), if each iteration is stopped at the first encounter of a local maximum for which  $\lambda(x_s) d(x_s, \epsilon_s) > m$ .

III. For a complete description of the  $D$ -optimal design, it is necessary to have its dispersion matrix (or information matrix) and the design itself.

$$\epsilon = \left\{ \begin{matrix} \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n \\ \hat{p}_1, \hat{p}_2, \dots, \hat{p}_n \end{matrix} \right\} \quad (2.6.6)$$

In numerically constructing the design  $\hat{\epsilon}$  we obtain the design  $\epsilon_s$ , which can be made arbitrarily close to  $\hat{\epsilon}$  but in general is usually distinct from it (we can take any large but finite number of iterations). This distinction will consist in the following:

1.  $(\hat{x}_i - x_{si})'(\hat{x}_i - x_{si}) = \xi_i$ , (2.6.7)

where  $\xi_i$  is some small positive number,

2.  $|\hat{p}_i - p_{si}| = \pi_i$ , (2.6.8)

where  $\pi_i$  is a small positive number.

3. The design  $\epsilon_s$  in comparison with the design  $\hat{\epsilon}$  will have constructed points

$$x_{s(n+1)}, x_{s(n+2)}, \dots, x_{s(n+l)} \quad (2.6.9)$$

with small weights

$$\gamma \geq p_{s(n+1)} \geq p_{s(n+2)} \geq \dots \geq p_{s(n+l)}. \quad (2.6.10)$$

4. Instead of one point  $x_{si}$  close to  $\hat{x}_i$  the design  $\epsilon_s$  will have a collection of points  $x_{si_1}, x_{si_2}, \dots, x_{si_l}$ , each of which is close to  $\hat{x}_i$ :

$$(\hat{x}_i - x_{si_k})'(\hat{x}_i - x_{si_k}) \leq \xi_i \quad (k = 1, 2, \dots, l), \quad (2.6.11)$$

and the sum of their weights will be close to  $\hat{p}_i$ :

$$\left| \sum_{k=1}^l p_{si_k} - \hat{p}_i \right| = \pi_i.$$

The above enumeration for a particular  $D$ -optimal design can be observed in Fig. 9C, where 15 iterations are presented for the regression function and the efficiency function considered in the example of Section 2.5.

Since designs with a large number of points are undesirable, instead of the design  $\epsilon_s$  it is usually more suitable to consider some approximation to it, in particular:

1. points with small weights not being attracted to any of the groups (2.6.11) are discarded;
2. points being attracted to one of the groups (2.6.11) are added to the group according to the rule

$$s_{si} = \hat{p}_{si}^{-1} \sum_{k=1}^l x_{si_k} p_{si_k}; \quad p_{si} = \sum_{k=1}^l p_{si_k}. \quad (2.6.12)$$

The design thus formed is verified to be "close" to a  $D$  optimal design. If the design is "close" to the optimal [e.g., satisfies (2.5.21)] then the computation is stopped. If not, the iterative process is continued. An example of such a rounded-off design is presented in Fig. 9D. In Fig. 9E the standard deviation for the design  $\epsilon_{11}$  and its

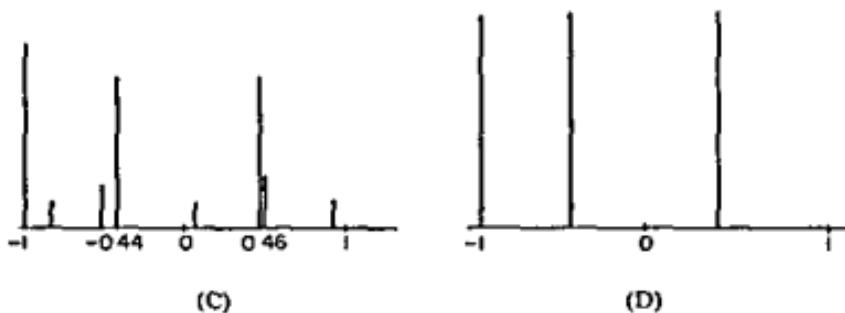


Fig. 9C Distribution of resources for the design  $\epsilon_{11}$

Fig. 9D The rounded-off design  $\epsilon_{11}$

Fig. 9E Dependence of  $\lambda(x) d(x, \epsilon)$  on  $x$ . The solid line corresponds to the design  $\epsilon_{11}$ , the dashed line to the design  $\epsilon_{11}$ .

rounded-off design is presented. The rounded-off design  $\epsilon_{11}$  is closer to the  $D$  optimal design than the design  $\epsilon_{11}$ . Generally speaking, this phenomenon is not accidental. Indeed those points with small weights [cf. (2.6.9) and (2.6.10)], and some distributions of points near the points of the  $D$  optimal design, are inherited from the nonoptimal design  $\epsilon_0$ . These, as a rule, are liquidated by the rounded-off design  $\epsilon_s$ .

IV. For large dimensions of the space of control variables the efficiency of the iterative method of constructing  $D$ -optimal designs can be significantly improved if from some general considerations it is possible to reduce, in comparison with the domain  $X$  of possible observations, the set of points which are studied in seeking  $\max_x \lambda(x) d(x, \epsilon_s)$ . In some cases the restriction of the region under investigation can be made from the beginning of the computation; in other cases such a reduction is possible in the process of computation.

EXAMPLE. We consider the linear two-dimensional regression

$$\eta(x, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2, \quad \lambda(x) \equiv 1, \quad x \in X.$$

The region of possible observations is presented in Fig. 10. Since the dispersion of the estimate of the response surface  $d(x, \epsilon)$  is a positive-definite quadratic form in the space  $x_1, x_2$  (cf. Example 1, Section 2.2), then it is obvious that the maximum value of  $d(x, \epsilon)$  must be obtained on the boundary of the region.

In connection with this, the function  $d(x, \epsilon_s)$  is investigated only on curves bounding the domain  $X$ . The spectrum of the rounded-off design  $\tilde{\epsilon}_3$  is plotted on the graph of the curve. The allocation is equal to  $p_1 = p_2 = p_3 = \frac{1}{3}$ . The maximum value of  $d(x, \tilde{\epsilon}_3)$  is 3.01. Recall that  $\max_x d(x, \tilde{\epsilon}) = 3$ .

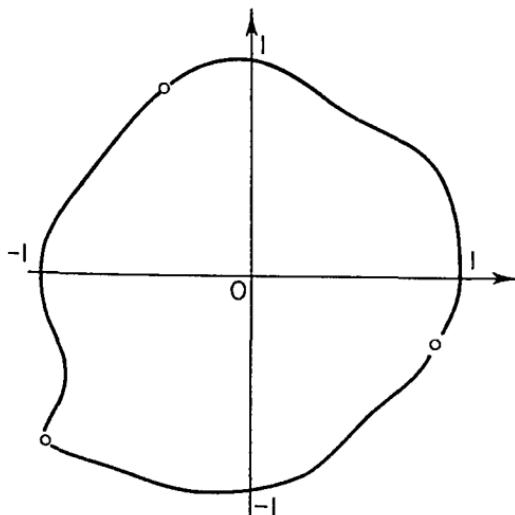


Fig. 10. An arbitrary region  $X$ . The circles indicate points corresponding to the spectrum of the design for linear regression.

V. In Part III of this section it was noted that the design  $\epsilon_s$  usually contains a significantly larger number of points than the design  $\epsilon$ . There a procedure was presented permitting us to cut down the number of points of the design  $\epsilon_s$  by means of replacing it by a rounded-off design. Since the  $D$ -optimal design contains at most  $m(m+1)/2$  points, keeping the designs  $\epsilon_s$  for computation on an electronic computing machine requires a large amount of memory. In the simplest case, when each memory cell contains a single number, the amount of necessary cells must be, in the best case (the design has a minimum number of points  $m$ )

$$S = km + m \quad (2.6.13)$$

where  $k$  is the dimension of the space of control variables. Recall that it is necessary to keep the coordinates of the point  $x_{sj}$  ( $j = 1, 2, \dots, k$ ) and weights  $p_{sj}$ .

In the case when the space of control variables has large dimension and there are a large number of unknown parameters the inconvenience of such a method of determining the characteristics of the  $D$ -optimal design is obvious. In this case there are many effective modifications of the iterative method considered on page 102.

After each  $s$ th iteration only the dispersion matrix  $D(\epsilon_s)$  is kept. The computations of the third step of each iteration (Section 2.5) are omitted, since there is no need for them in determining the elements of  $D(\epsilon_s)$ . Correspondingly, the necessity of keeping the numerical characteristics of the design  $\epsilon_s$  is omitted. The iterative procedure, as earlier, continues so long as the given conditions for ending the computation are not satisfied. For further computations only the elements of the matrix  $D(\epsilon_s)$  are necessary.

After the iterative procedure is ended [e.g., one of the conditions (2.5.21) holds], all maxima  $x_i$  of the function  $\lambda(x) d(x, \epsilon_s)$  are sought for which

$$\lambda(x_i) d(x_i, \epsilon_s) \cong m \quad (2.6.14)$$

Such points will be no less than  $m$ . By definition

$$\sum_{i=1}^n p_{si} f_a(x_{si}) f_b(x_{si}) = M_{ab}(\epsilon_s) = M_{ab}(\epsilon_s), \quad (2.6.15)$$

$$D(\epsilon_s) = M^{-1}(\epsilon_s)$$

We introduce the following matrices:

$$p' = \| p_1, p_2, \dots, p_n \|, \quad n = m(m+1)/2,$$

$$\mathcal{M}'(\epsilon) = \| M_{11}(\epsilon), M_{12}(\epsilon), \dots, M_{\alpha\beta}(\epsilon), \dots, M_{mm}(\epsilon) \|, \quad m \geq \beta \geq \alpha \geq 1,$$

$$\Phi = \| \mathcal{M}[\epsilon(x_1)], \mathcal{M}[\epsilon(x_2)], \dots, \mathcal{M}[\epsilon(x_n)] \|,$$

where the spectrum of  $\epsilon(x_i)$  ( $i = 1, 2, \dots, n$ ) contains one point.

If  $n < m(m+1)/2$ , then from the matrix  $\Phi$  one of the nondegenerate matrices  $\Phi_n$  of dimension  $n \times n$  is extracted. The necessary weights can be defined by solving the system

$$\Phi_n p = \mathcal{M}_n(\epsilon_s), \quad (2.6.16)$$

when  $\mathcal{M}_n(\epsilon_s)$  is the part of the vector  $\mathcal{M}(\epsilon_s)$ , corresponding to the matrix  $\Phi_n$ . As was shown in Theorem 2.5.3

$$\lim_{s \rightarrow \infty} D(\epsilon_s) = D(\hat{\epsilon}), \quad (2.6.17)$$

$$\lim_{s \rightarrow \infty} \lambda(x) d(x, \epsilon_s) = \lambda(x) d(x, \hat{\epsilon}).$$

From (2.6.16) and (2.6.17) the following set of equalities obviously follows:

$$\lim_{s \rightarrow \infty} x_{si} = \hat{x}_i, \quad (i = 1, 2, \dots, n), \quad (2.6.18)$$

where  $x_{si}$  corresponds to the  $i$ th maximum of the function  $\lambda(x) d(x, \epsilon_s)$ ;

$$\lim_{s \rightarrow \infty} \Phi(x_{s1}, x_{s2}, \dots, x_{sn}) = \Phi(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n), \quad (2.6.19)$$

$$\lim_{s \rightarrow \infty} p_s = \hat{p} \quad (2.6.20)$$

Here  $\hat{x}_i$  is the coordinate of the point belonging to the spectrum of the  $D$ -optimal design  $\hat{\epsilon}$  and  $\hat{p}$  is the corresponding vector of allocations.

If  $p_i$  is small or zero,  $p_{si}$  can take on a negative value with small modulus. For definiteness, let  $n = m(m+1)/2$ . In this case the phenomenon of negative weights is explained by the fact that the information matrix of the design  $\epsilon_s$ , in distinction from the information matrix of the design  $\hat{\epsilon}$ , is an interior point of the set of all possible information matrices for the given closed region  $X$  and the regression function  $\eta(x, \theta)$  and the efficiency function  $\lambda(x)$ . It follows by Theorem 2.1 that the design  $\epsilon_s$  (it is not exactly  $D$ -optimal) can be represented in the form of a linear combination with positive coefficients [of no

more than  $m(m+1)/2 + 1$  points] of the designs  $\epsilon(x_i)$  [recall that for a  $D$  optimal design this bound equals  $m(m+1)/2$ ]

But the system (2.6.16) has no more than  $m(m+1)/2$  equations, which is one less than that required for a sufficient condition in Theorem 2.1. Therefore in some cases any of the  $p_{st}$  can appear to be less than zero.

**VI** The convergence of the iterative procedure considered in Section 2.5 to the  $D$  optimal was proved for the case when  $\alpha_s$  was chosen according to (2.5.16). By analogy to Theorem 2.5.3 it is possible to prove that the iterative process presented converges to the  $D$  optimal for any sequence  $\alpha_s$  satisfying the conditions

$$\sum_{s=1}^{\infty} \alpha_s = \infty \quad \lim_{s \rightarrow \infty} \alpha_s = 0 \quad (2.6.21)$$

One such sequence is the sequence  $\alpha_s \sim s^{-1}$ .

The search for an optimal sequence  $\{\alpha_s\}$  (i.e. such that a given accuracy is attained after the smallest number of iterations) has not been successful. Evidently for each set of functions  $\eta(x, \theta)$  and  $\lambda(x)$  there exists a separate optimal sequence.

The practice of numerical construction of  $D$  optimal designs shows that the most successful sequence  $\{\alpha_s\}$  is defined by (2.5.16) and the sequence  $\{\alpha_s\}$  constructed in the following manner:

1. Choose some  $\alpha_0$ , for which it is expected to decrease the determinant  $|D(\epsilon_0)|$ . The iteration with the specified  $\alpha_0$  is continued as long as the determinant  $|D(\epsilon_s)|$  is observed to decrease.

2. As soon as it appears that

$$|D(\epsilon_{s-1})| \geq |D(\epsilon_s)| \quad (2.6.22)$$

$\alpha_0$  is decreased, for example, it is possible to divide it by two.

3. Afterward the iteration is repeated with  $\alpha_0/2$  until (2.6.22) holds, then a reduction to  $\alpha_0/4$  is made and so on.

## 2.7. Truncated $D$ Optimal Designs

I. In many experimental investigations the following situation presents itself. The observed quantities depend, for example, on unknown parameters according to the rule

$$E(y|x) = \eta(x|\theta) = \theta f(x) \quad (2.7.1)$$

Among the parameters  $\theta_\alpha$  ( $\alpha = 1, 2, \dots, m$ ) there are several mixing parameters  $\theta_\alpha$  ( $\alpha = l+1, l+2, \dots, m$ ), which exert influence on the investigated variables  $y$ , but the experimenter is not particularly interested in them. In this case, the  $D$ -optimal designs considered earlier obviously do not reflect the needs of the experimenter, and it follows that they are inappropriate. The experiments must satisfy different criteria of optimality immediately related to the accuracy of the determination of the estimates only of those parameters which interest the investigator ("necessary parameters"). We will call such criteria "truncated."

II. The design  $\hat{\epsilon}$  is called a truncated  $D$ -optimal design if it maximizes the value of the determinant of the submatrix  $D_{ll}$  of the dispersion matrix  $D(\epsilon)$  of estimates of the unknown parameters  $\theta$ :

$$|D_{ll}(\hat{\epsilon})| = \min_{\epsilon} |D_{ll}(\epsilon)|. \quad (2.7.2)$$

The properties of the truncated  $D$ -optimal designs, in the large, resemble the properties of the  $D$ -optimal designs.

In particular, a theorem analogous to Theorem 2.2.1 holds and was first proved by Kiefer [20].

**Theorem 2.7.1.** *The following assertions:*

- (1) *the design  $\hat{\epsilon}$  minimizes  $|D_{ll}(\epsilon)|$ ;*
- (2) *the design  $\hat{\epsilon}$  minimizes  $\max_x \lambda(x) d(x, l, \epsilon)$ , where*

$$d(x, l, \epsilon) = f'(x) \begin{vmatrix} D_{ll} & D_{lk} \\ D_{kl} & D_{kk} - M_{kk}^{-1} \end{vmatrix} f(x),$$

- $l+k=m$ , and  $M_{kk}$  is a submatrix of the information matrix  $M(\epsilon)$ ;
- (3)  $\max_x \lambda(x) d(x, l, \epsilon) = l$

*are equivalent. Any linear combinations of designs satisfying conditions 1-3 also satisfies these conditions.<sup>4</sup>*

*Proof.* (1) We will show that (2) follows from (1). We consider the design  $\tilde{\epsilon} = (1-\alpha)\hat{\epsilon} + \alpha\epsilon$ . By the definition of the design  $\hat{\epsilon}$

$$(\partial/\partial\alpha) \log |D_{ll}(\tilde{\epsilon})| \Big|_{\alpha=0} \geq 0. \quad (2.7.3)$$

<sup>4</sup> This theorem seems to require the condition  $|M_{kk}| \neq 0$ , cf. C. L. Atwood, Optimal and efficient designs of experiments. *Ann. Math. Statist.* 40, 1570-1602 (1969).—TRANS.

We compute the derivative on the left side of the preceding inequality. By (1.1.17)

$$|D_{tt}(\tilde{\epsilon})| = |M_{kk}(\tilde{\epsilon})| |M(\tilde{\epsilon})|,$$

where  $M(\tilde{\epsilon}) = (1 - \alpha) M(\tilde{\epsilon}) + \alpha M(\epsilon)$ . Using the formulas introduced in Part VI of Section 1.1 we obtain

$$\begin{aligned}\frac{\partial}{\partial x} \log \frac{|M_{kk}(\tilde{\epsilon})|}{|M(\tilde{\epsilon})|} &= \frac{|M(\tilde{\epsilon})| (\tilde{\epsilon}' \tilde{\epsilon}_x) |M_{kk}(\epsilon)| - |M_{kk}(\tilde{\epsilon})| (\tilde{\epsilon}' \tilde{\epsilon}_x) |M(\epsilon)|}{|M_{kk}(\epsilon)| |M(\epsilon)|} \\ &= \text{Tr } M_{kk}^{-1}(\epsilon) \frac{\partial}{\partial x} M_{kk}(\epsilon) - \text{Tr } M^{-1}(\tilde{\epsilon}) \frac{\partial}{\partial x} M(\epsilon),\end{aligned}$$

or

$$\begin{aligned}(\tilde{\epsilon}' \tilde{\epsilon}_x) \log |D_{tt}(\tilde{\epsilon})| \Big|_{\alpha=0} &= \text{Tr } M_{kk}^{-1}(\tilde{\epsilon}) M_{kk}(\epsilon) - \text{Tr } M_{kk}^{-1}(\tilde{\epsilon}) M_{kk}(\tilde{\epsilon}) \\ &\quad - \text{Tr } M^{-1}(\tilde{\epsilon}) M(\epsilon) + \text{Tr } M^{-1}(\tilde{\epsilon}) M(\tilde{\epsilon}) \\ &= l - \text{Tr } d(\tilde{\epsilon}) M(\epsilon),\end{aligned}\tag{2.7.4}$$

where

$$d = \left\| \begin{array}{cc} D_{tt} & D_{tk} \\ D_{kt} & D_{kk} - M_{kk}^{-1} \end{array} \right\|$$

We choose as an  $\epsilon$  design the allocation at a single point  $x$ . Then from (2.7.3) and (2.7.4) we obtain

$$\text{Tr } d(\tilde{\epsilon}) \lambda(x) f(x) f'(x) = \lambda(x) d(x, l, \tilde{\epsilon}) \leq l\tag{2.7.5}$$

On the other hand, for any design  $\epsilon$  (as in the case of  $D$ -optimal designs, by Theorem 2.1.1, it is possible to limit ourselves to designs with discrete finite spectrum)

$$\begin{aligned}\sum_{i=1}^n p_i \lambda(x_i) d(x_i, l, \epsilon) &= \sum_{i=1}^n p_i \lambda(x_i) f(x_i) d(\epsilon) f(x_i) \\ &= \text{Tr } M(\epsilon) d(\epsilon) = \text{Tr } M(\epsilon) M^{-1}(\epsilon) \\ &\quad - \text{Tr } M(\epsilon) \left\| \begin{array}{cc} 0 & 0 \\ 0 & M_{kk}^{-1}(\epsilon) \end{array} \right\| \\ &= \text{Tr } I_m - \text{Tr } I_k = l,\end{aligned}\tag{2.7.6}$$

or since  $\sum_{i=1}^n p_i = 1$ ,

$$\max_x \lambda(x) d(x, l, \epsilon) \geq \max_i \lambda(x_i) d(x_i, l, \epsilon) \geq l. \quad (2.7.7)$$

From (2.7.5) and (2.7.7) it follows that the truncated D-optimal design minimizes  $\max_x \lambda(x) d(x, l, \epsilon)$ .

(2) We show that (1) follows from (2). We will assume that the design  $\hat{\epsilon}$  does not satisfy (1). We will construct a design  $\tilde{\epsilon} = (1 - \alpha) + \alpha\hat{\epsilon}$ , where  $\hat{\epsilon}$  is one of the designs minimizing  $|D_u(\epsilon)|$ . The matrix  $D_u(\tilde{\epsilon})$  can be written in the form (cf. Theorem 1.1.4)

$$M_u(\tilde{\epsilon}) = D_u^{-1}(\tilde{\epsilon}) = M_{ll}(\tilde{\epsilon}) - M_{lk}(\tilde{\epsilon}) M_{kk}^{-1}(\tilde{\epsilon}) M_{kl}(\tilde{\epsilon}). \quad (2.7.8)$$

From (2.7.8) and Theorem 1.1.13 it follows that  $M_{ll}(\tilde{\epsilon}) - (1 - \alpha) \times M_{ll}(\hat{\epsilon}) - \alpha M_{ll}(\hat{\epsilon}) \geq 0$ . From this and Theorem 1.1.14

$$\begin{aligned} \log |M_u(\tilde{\epsilon})| &\geq \log |(1 - \alpha) M_u(\hat{\epsilon}) + \alpha M_u(\hat{\epsilon})| \\ &\geq (1 - \alpha) \log |M_u(\hat{\epsilon})| + \alpha \log |M_u(\hat{\epsilon})|. \end{aligned} \quad (2.7.9)$$

Inequality (2.7.9) says that the function  $\log |M_u(\tilde{\epsilon})|$  is concave.

Considering the definition of the design  $\hat{\epsilon}$  we obtain from (2.7.9)

$$\log |D_u(\tilde{\epsilon})| \leq (1 - \alpha) \log |D_u(\hat{\epsilon})| + \alpha \log |D_u(\hat{\epsilon})| < \log |D_u(\hat{\epsilon})|. \quad (2.7.10)$$

Since (2.7.10) holds for any  $\alpha > 0$ , this implies that for the design  $\tilde{\epsilon}$

$$(\partial/\partial\alpha) \log |D_u(\tilde{\epsilon})| \Big|_{\alpha=0} < 0. \quad (2.7.11)$$

On the other hand,

$$(\partial/\partial\alpha) \log |D_u(\tilde{\epsilon})| \Big|_{\alpha=0} = l - \text{Tr } A(\hat{\epsilon}) M(\hat{\epsilon}). \quad (2.7.12)$$

We write (2.7.12) in the form

$$(\partial/\partial\alpha) \log |D_u(\tilde{\epsilon})| \Big|_{\alpha=0} = l - \sum_{i=1}^n p_i \lambda(x_i) d(x_i, l, \hat{\epsilon}), \quad (2.7.13)$$

where the sum extends over the spectrum of the design  $\hat{\epsilon}$ . From (2.7.13) and the definition of the design  $\hat{\epsilon}$ , it follows that

$$(\partial/\partial\alpha) \log |D_u(\epsilon)| \Big|_{\alpha=0} \geq l - \sum_{i=1}^n p_i l = 0 \quad (2.7.14)$$

The contradiction obtained [cf (2.7.11) and (2.7.14)] proves our assertion

(3) The equivalence of assertions (1) and (3) and (2) and (3) immediately follows from the equivalence of assertions (1) and (2) and inequality (2.7.7)

(4) The concluding assertion immediately follows from the concavity of the function  $\log |\mathcal{M}_D(\epsilon)|$

The theorem is proved

*Remark 1* In the proof it was assumed that  $|M(\hat{\epsilon})| \neq 0$ , but sometimes truncated D-optimal designs have singular information matrices. In these cases the matrix  $D$  is defined by

$$D = \tilde{D} L D_{II}^{-1} L \tilde{D}$$

where

$$I = \begin{pmatrix} I_1 & 0 \end{pmatrix},$$

$$\tilde{D} = U \begin{pmatrix} A_{nn}^{-1} & 0 \\ 0 & 0 \end{pmatrix} U \quad (2.7.15)$$

the matrices  $U$  and  $A_{nn}$  are determined from the relationship  $UU^T = I$  and

$$UM(\hat{\epsilon})U = A - \begin{pmatrix} A_{nn} & 0 \\ 0 & 0 \end{pmatrix}$$

$A_{ss} = \lambda_s \delta_{ss}$ ,  $\lambda_s > 0$  for  $s \leq n$ ,  $n = \text{rank } M(\hat{\epsilon})$ . We note that  $D_{II} = \tilde{D}_{II}$ . The matrix  $D_{II}$  may also be computed making use of passage to the limit (cf Section 1.3)

**Corollary 1** At the points of the spectrum of the truncated D-optimal design  $\hat{\epsilon}$  the quantity  $\lambda(x) d(x, l, \hat{\epsilon})$  attains its maximal value  $l$

The reader can verify without difficulty the given assertion by considerations analogous to those used for the proof of Corollary 1, of Theorem 2.2.3

In Section 2.2 it was shown that a D-optimal design containing no more than  $m(m+1)/2$  points can always be found. For truncated D-optimal designs, this bound can be lowered

First, we note that the matrix  $D_{II}(\epsilon)$  is fully characterized [cf (2.7.8)] by  $n_0 = [l(l+1)/2] + lk = (l/2)(2k+l+1)$  elements of the infor-

mation matrix. From this it is not difficult to see (cf. the concluding part of Theorem 2.1.1) that for any design  $\epsilon$  with a given matrix  $D_{ll}(\epsilon)$  a design  $\epsilon_0$  can be found, the spectrum of which contains no more than  $n_0 + 1$  points, and  $D_{ll}(\epsilon) = D_{ll}(\epsilon_0)$ .

For any truncated  $D$ -optimal design  $\tilde{\epsilon}_0$  a design  $\tilde{\epsilon}$  can always be found, the spectrum of which contains  $n_0$  points and  $D_{ll}(\tilde{\epsilon}) = D_{ll}(\tilde{\epsilon}_0)$ . The last assertion is proved in complete analogy with Theorem 2.2.3.

**III.** For numerical construction of truncated  $D$ -optimal designs it is possible to use the iterative procedure considered in Sections 2.5 and 2.6 for  $D$ -optimal designs. In this case, instead of  $\max_x \lambda(x) d(x, \epsilon_s)$  we now seek  $\max_x \lambda(x) d(x, l, \epsilon_s)$ . The step  $\alpha_s$  can be chosen corresponding to

$$\max_{\alpha} (1 - \alpha)^{l-m} \frac{1 - \alpha - \alpha \lambda(x_s) d(x_s, l, \epsilon_s)}{1 - \alpha + \alpha \lambda(x_s) d(x_s, \epsilon_s)}. \quad (2.7.16)$$

**Theorem 2.7.2.** *The iterative procedure presented above converges;  $\lim_{s \rightarrow \infty} |D_{ll}(\epsilon_s)| = |D_{ll}(\tilde{\epsilon})|$ . In this case, if the design  $\tilde{\epsilon}$  is nondegenerate, then*

$$|D_{ll}(\tilde{\epsilon})| = \min_{\epsilon} |D_{ll}(\epsilon)|. \quad (2.7.17)$$

The proof is analogous to the proof of Theorem 2.5.3 and can be carried out by the reader independently.

It is to be pointed out that the design  $\tilde{\epsilon}$  is always better than the initial design  $\epsilon_0$  (if the latter does not coincide with the truncated  $D$ -optimal design):  $|D_{ll}(\tilde{\epsilon})| < |D_{ll}(\epsilon_0)|$ . If the initial approximation  $\epsilon_0$  is chosen sufficiently close to the optimal design, equality (2.7.17) holds even for degenerate designs  $\tilde{\epsilon}$ .

The possibility of convergence of the iterative procedure for  $\alpha_s (s = 0, 1, \dots)$ , chosen based on (2.7.16) when there exist degenerate optimal designs, to the design distinct from the optimal necessitates that the expression  $\lambda(x_s) d(x_s, \epsilon_s)$  entering into the denominator of (2.7.16) grow rapidly. This evokes, in turn, an excessively rapid decrease of  $\alpha_s$ .

If the sequence  $\{\alpha_s\}$  is chosen such that  $\sum_{s=1}^{\infty} \alpha_s = \infty$ ,  $\lim_{s \rightarrow \infty} \alpha_s = 0$ , or is chosen such that  $\alpha_s$  decreases in  $\gamma > 1$ , as soon as the inequality holds

$$|D_{ll}(\epsilon_{s+1})| \geq |D_{ll}(\epsilon_s)|,$$

then it is possible to show that Theorem 2.7.2 will also hold for the corresponding iterative procedure.

In carrying out the methods considered for numerical construction of truncated  $D$ -optimal designs one should consider the recommendations introduced in Section 2.6.

Since the properties of  $D$ -optimal and truncated  $D$ -optimal designs are similar to one another, then, where this cannot distort the sense of the text, we will unify these and other designs under the general title  *$D$ -optimal designs*.

## 2.8. Nonlinear Parametrization of a Response Surface.

### Local $D$ -Optimal Designs

We will assume that the function  $\eta(x, \theta)$  is a nonlinear function of the parameter  $\theta$  and that the effectiveness  $\lambda(x)$  of the experiment at which the measurements of the quantity  $y$  are taken is known.

As was shown in Section 1.4, from the practical viewpoint best quasi-linear estimates  $\hat{\theta}$  are most useful when their dispersion matrix is given by

$$D(\hat{\theta}, \epsilon) \simeq M^{-1}(\theta_0, \epsilon), \quad (2.8.1)$$

where  $M(\theta_0, \epsilon) = \sum_{i=1}^n n_i \lambda(x_i) f(x_i) f'(x_i)$  and  $n_i$  is the number of measurements taken at the point  $x_i$  ( $i = 1, 2, \dots, n$ ) [the remaining notation is given in the clarification to (1.4.5)].

It is evident that the matrix  $M$  observes the same properties as the Fisher information matrix, defined for the case of linear parametrization of a response surface. In particular, Theorems 2.1.2 and 2.2.4 are valid for it.

From Theorem 2.2.4, for example, it follows that when (2.8.1) holds [cf. also remarks to formulas (1.4.7) and (1.4.8)] for the design  $\epsilon$  of minimal determinant  $|D(\theta_0, \epsilon)|$ , it is always possible to find a design with the same value of the determinant containing no more than  $m(m+1)/2$  points. In the frame of approximation (2.8.1), the remaining results of Section 2.2 and the results of Sections 2.5–2.8 are valid. Since the matrix depends on the true values of the sought parameters  $\theta_0$  (cf. Section 1.4), then, in general, we cannot *a priori* construct  $D$ -optimal designs. The maximum possible for us is the construction of optimal designs under the assumption that the true values of the sought parameters are given to us, i.e.,  $\theta_0 = \theta$ . This

indicates that the  $D$ -optimal design for nonlinear parametrization is a function of the unknown parameters  $\theta$ . In order to emphasize this fact, the design  $\dot{\epsilon}(\theta_0)$  is called a local  $D$ -optimal design [31]. The analytic construction of these designs is possible only in the simplest cases.

Several simple situations are those cases when it is possible *a priori* to partition the region  $\Omega$  into subregions, one of which obviously (or with sufficiently large probability) contains the point  $\theta_0$ . If in this subregion the values of the derivatives  $f_\alpha(x) = (\partial/\partial\theta_\alpha)\eta(x, \theta)$  ( $\alpha = 1, 2, \dots, m$ ) change very little, then the problem of design becomes equivalent to the problem of design for linear parametrization of the response surface.

However, it is to be pointed out that the partition is possible only in the presence of some prior information. As will be shown later (see Chapter 4), in these cases a sequence of designs considering prior information is more effective.

**EXAMPLE [31].** We assume that  $p_x = e^{-\theta x}$  ( $\theta > 0, x \geq 0$ ) is the probability with which a microorganism lives after being subjected to the influence of some substance, the dose of which is equal to  $x$ .

The usual experiment is conducted in the following way. Some number  $N$  of microorganisms are chosen. These are divided into  $n$  groups with  $n_i$  ( $i = 1, 2, \dots, n$ ) pieces in each. Each group is subjected to the influence of the investigated substance with a dose  $x_i$  ( $i = 1, 2, \dots, n$ ). The number  $\tilde{n}_i$  of surviving microorganisms is counted in each group and a regression analysis is carried out for the determination of the estimate of the parameter of activeness  $\theta$  [we note that  $E(\tilde{p}_i | x_i) = e^{-\theta x_i}$ , where  $\tilde{p}_i = \tilde{n}_i/n_i$ ].

Since the number of unknown parameters  $m$  is 1, by the results presented earlier we can find an optimal design consisting of a single point  $\dot{x}$ . It follows that the division into groups is not advantageous.

For seeking the optimal point  $\dot{x}$ , it is necessary to know the effectiveness of the designed experiment. It is not difficult to verify that  $D(\tilde{p}_i) = p_i(1 - p_i)/n_i$ . Therefore  $\lambda(x_i) n_i = D^{-1}(\tilde{p}_i)$  and  $\lambda^{-1}(x_i) = p_i(1 - p_i)$ ,  $x_i > 0$ . Considering that  $f(x) = \partial e^{-\theta x}/\partial\theta = -xe^{-\theta x}$  and the observations in the optimal design must be conducted at a single point, we compute the matrix  $M$  (consisting here of one element) under the assumption that  $\theta$  is given:

$$NM(\epsilon) = N\lambda(x)f^2(x) = N[x^2e^{-\theta x}/(1 - e^{-\theta x})], \quad x > 0,$$

and

$$NM(\epsilon) = 0 \quad x = 0$$

The local  $D$  optimal design then says that all  $N$  microorganisms must be subjected to the same dose  $x$  of the substance where  $x$  is determined from the equation

$$(\partial/\partial x)[x^2 e^{-\theta x}/(1 - e^{-\theta x})] = 0$$

or

$$2e^{-\theta x} + \theta x = 2$$

and is approximately  $1.6\theta^{-1}$ . In this manner the given locally  $D$  optimal design depends critically on the true value of the parameter of the activeness of the investigated substance.

## 2.9 Linear Criteria of Optimality

I Let the designs of experiments be compared according to the rule  $\epsilon_1 > \epsilon_2$  if

$$L[D(\epsilon_1)] < L[D(\epsilon_2)] \quad (2.9.1)$$

The functional  $L$  sets in correspondence to each matrix  $D(\epsilon)$  some scalar quantity. We will assume that this functional is linear that is

$$L(A + B) = L(A) + L(B) \quad (2.9.2)$$

$$L(cA) = cL(A) \quad (2.9.3)$$

We require also that the inequality

$$L(A) \geq 0 \quad (2.9.4)$$

be satisfied for all positive semidefinite matrices  $A$ . For brevity we will call the design  $\epsilon$  *linear optimal* if

$$L[D(\epsilon)] = m n L[\epsilon] \quad (2.9.5)$$

II We consider basic properties of linear optimal designs

**Theorem 2.9.1** For any linear optimal design  $\epsilon$  one can always find a design  $\epsilon_0$  the spectrum of which consists of no more than  $n_0$  points and

such that

$$L[D(\hat{\epsilon}_0)] = L[D(\hat{\epsilon})].$$

Here  $n_0$  is the number of elements of the information matrix on which the quantity  $L[D(\epsilon)]$  depends [recall that the information matrix is completely characterized by its  $m(m + 1)/2$  elements].

*Proof.* We identify the  $n_0$  elements of the information matrix with elements of  $n_0$ -dimensional Euclidean space. The collection of all possible designs in this space will correspond to some closed set  $\mathcal{E}$  (cf. Theorem 2.1.1), appearing as a convex combination of the set of points, corresponding to the set of all single-point designs.

We show that the point corresponding to  $M(\hat{\epsilon})$  is a boundary point of this set.

We assume the contrary. Let the given point be interior. Then by the definition of an interior point of a set (cf. Section 2.1), there is an  $\alpha > 0$  such that the point corresponding to  $M(\epsilon) = (1 + \alpha) M(\hat{\epsilon})$  also belongs to the set  $\mathcal{E}$ . From this and (2.9.3) it follows that

$$L[M^{-1}(\epsilon)] = (1 + \alpha)^{-1} L[M^{-1}(\hat{\epsilon})] < L[M^{-1}(\hat{\epsilon})]. \quad (2.9.6)$$

Inequality (2.9.6) contradicts the definition of the design  $\hat{\epsilon}$ . It follows that  $M(\hat{\epsilon})$  is a boundary point of the set  $\mathcal{E}$ .

From this and the theorem of Carathéodory it follows that the matrix  $M(\hat{\epsilon})$  can be represented as a linear combination  $\sum_{i=1}^{n_0} p_i M[\epsilon(x_i)]$ , where  $M[\epsilon(x_i)]$  corresponds to the single-point design  $\epsilon(x_i)$ . Therefore any design  $\hat{\epsilon}$  can be replaced by a linear combination of  $n_0$  single-point designs. The theorem is proved.

Theorem 2.9.1 has an important practical significance, since it permits us to limit our considerations to designs having finite spectrum. In what follows if it is not otherwise stated we will consider only such designs.

### Lemma 2.9.1

- (1) *The function  $L[M^{-1}(\epsilon)]$  is a convex function of  $M(\epsilon)$ .*
- (2) *If instead of (2.9.4) the strengthened condition*

$$L(A) > 0 \quad (2.9.7)$$

*for any positive-semidefinite matrix is used, then the function  $L[M^{-1}(\epsilon)]$  is a strictly convex function of  $M(\epsilon)$  on the set of continuous designs.*

*Proof* From Theorem 2.1.2 it follows that the collection of information matrices  $M(\epsilon)$  is a closed convex set. From this and the definition of a convex function, for the proof of the first part of the lemma, it is sufficient to show that

$$L\{[(1-\alpha)M(\epsilon_1) + \alpha M(\epsilon_2)]^{-1}\} \leq (1-\alpha)L[M^{-1}(\epsilon_1)] + \alpha L[M^{-1}(\epsilon_2)] \quad (2.9.8)$$

for any  $M(\epsilon_1)$  and  $M(\epsilon_2)$ . The inequality (2.9.8) follows from (2.9.2) and (2.9.4), and the fact that (cf. Theorem 1.1.12)

$$(1-\alpha)M^{-1}(\epsilon_1) + \alpha M^{-1}(\epsilon_2) \geq [(1-\alpha)M(\epsilon_1) + \alpha M(\epsilon_2)]^{-1} \quad (2.9.9)$$

For the proof of the second part it is necessary to show that

$$L\{[(1-\alpha)M(\epsilon_1) + \alpha M(\epsilon_2)]^{-1}\} < (1-\alpha)L[M^{-1}(\epsilon_1)] + \alpha L[M^{-1}(\epsilon_2)] \quad (2.9.10)$$

for  $M(\epsilon_1) \neq M(\epsilon_2)$ ,  $0 < \alpha < 1$ .

Since the equality in (2.9.9) holds only when  $M(\epsilon_1) = M(\epsilon_2)$  (cf. Theorem 1.1.12), then (2.9.10) follows from (2.9.7) and (2.9.8). The lemma is proved.

*Remark 1* Here and in what follows, in those cases where the information matrix is degenerate all assertions are interpreted according to Remark 1 to Theorem 2.7.1.

**Lemma 2.9.2.** Let  $\epsilon = (1-\alpha)\epsilon_1 + \alpha\epsilon_2$ , then

$$\partial D(\epsilon) / \partial \alpha = -D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) \quad (2.9.11)$$

and

$$\partial^2 D(\epsilon) / \partial \alpha^2 = 2D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) \geq 0 \quad (2.9.12)$$

*Proof* Since

$$D^{-1}(\epsilon) = M(\epsilon) = (1-\alpha)M(\epsilon_1) + \alpha M(\epsilon_2) \quad (2.9.13)$$

then from (1.1.13) and (2.9.13) it follows that

$$\begin{aligned} \partial D(\epsilon) / \partial \alpha &= -D(\epsilon)(\partial / \partial \alpha)[(1-\alpha)M(\epsilon_1) + \alpha M(\epsilon_2)] D(\epsilon) \\ &= -D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) \end{aligned}$$

Differentiating (2.9.11) with respect to  $\alpha$ , we obtain

$$\partial^2 D(\epsilon) / \partial \alpha^2 = 2D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) \quad (2.9.14)$$

We multiply (2.9.14) on the left by  $q'$  and on the right by  $q$ :

$$\begin{aligned} q'[\partial^2 D(\epsilon)/\partial \alpha^2] q &= 2q'D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon)[M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) q \\ &= l'D(\epsilon) l, \end{aligned} \quad (2.9.15)$$

where  $l = [M(\epsilon_2) - M(\epsilon_1)] D(\epsilon) q$ .

Since the matrix  $D(\epsilon)$  is positive semidefinite, then  $l'D(\epsilon) l \geq 0$ , and it follows that

$$q'[\partial^2 D(\epsilon)/\partial \alpha^2] q \geq 0,$$

that is, the matrix  $\partial^2 D(\epsilon)/\partial \alpha^2$  is a positive-semidefinite matrix. The lemma is proved.

**Theorem 2.9.2.** *The following assertions:*

- (1) *the design  $\hat{\epsilon}$  minimizes  $L[D(\epsilon)]$ ,*
- (2) *the design  $\hat{\epsilon}$  minimizes  $\max_x \lambda(x) L[D(\epsilon) f(x) f'(x) D(\epsilon)]$ ,*
- (3)  $\max_x \lambda(x) L[D(\hat{\epsilon}) f(x) f'(x) D(\hat{\epsilon})] = L[D(\hat{\epsilon})]$

*are equivalent. Any linear combination of designs satisfying (1)–(3), also satisfies (1)–(3).*

*Remark 2.* If the linear-optimal design has a singular matrix  $M(\hat{\epsilon})$ , then instead of the function  $\lambda(x) L[D(\epsilon) f(x) f'(x) D(\epsilon)]$  in the formulation of the theorem, the function  $\lambda(x) L[\tilde{D}(\epsilon) f(x) f'(x) \tilde{D}(x)]$  must be used, where  $\tilde{D}(\epsilon)$  is defined in the remarks to Theorem 2.7.1.

*Proof.* (1) We will show that (2) follows from (1). To do this, we consider the design  $\tilde{\epsilon} = (1 - \alpha)\hat{\epsilon} + \alpha\epsilon$ , where  $\epsilon$  is some arbitrary design. In view of the linearity of the functional  $L$

$$(\partial/\partial \alpha) L[D(\tilde{\epsilon})] = L[(\partial/\partial \alpha) D(\tilde{\epsilon})]. \quad (2.9.16)$$

From (2.9.16) and Lemma 2.9.2 we obtain

$$(\partial/\partial \alpha) L[D(\tilde{\epsilon})]_{\alpha=0} = L\{D(\hat{\epsilon})[M(\hat{\epsilon}) - M(\epsilon)] D(\hat{\epsilon})\}.$$

In view of the definition of the design  $\hat{\epsilon}$ , and the convexity of  $L[M^{-1}(\epsilon)]$  the following inequality must be satisfied:

$$(\partial/\partial \alpha) L[D(\tilde{\epsilon})] \Big|_{\alpha=0} \geq 0. \quad (2.9.17)$$

If the design  $\epsilon$  is concentrated at a single point  $x$ , then (2.9.17) takes on the form

$$(\partial/\partial\alpha)L[D(\epsilon)] = I[D(\hat{\epsilon})] - \lambda(x)L[D(\hat{\epsilon})f(x)f'(x)D(\hat{\epsilon})] \geq 0 \quad (2.9.18)$$

since  $M[\epsilon(x)] = \lambda(x)f(x)f'(x)$

On the other hand, for any design  $\epsilon$ ,

$$\begin{aligned} & \sum_{i=1}^n p_i \lambda(x_i) L[D(\epsilon)f(x_i)f'(x_i)D(\epsilon)] \\ &= L\left[D(\epsilon) \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i) D(\epsilon)\right] \\ &= L[D(\epsilon) M(\epsilon) D(\epsilon)] = I[D(\epsilon)] \end{aligned} \quad (2.9.19)$$

or since  $\sum_{i=1}^n p_i = 1$ ,

$$\begin{aligned} & \max_x \lambda(x) I[D(\epsilon)f(x)f'(x)D(\epsilon)] \\ & \geq \max_i \lambda(x_i) L[D(\epsilon)f(x_i)f'(x_i)D(\epsilon)] \\ & \geq L[D(\epsilon)] > L[D(\hat{\epsilon})] \end{aligned} \quad (2.9.20)$$

Comparing (2.9.18) and (2.9.20) it is not difficult to see that for a linear-optimal design,  $\max_x q(x, \hat{\epsilon}) = \max_x \lambda(x)L[D(\hat{\epsilon})f(x)f'(x)D(\hat{\epsilon})]$  attains the smallest possible value.

In further considerations, for brevity we call the design minimizing  $\max_x q(x, \epsilon)$  the minimax design.

(2) We show that (1) follows from (2). We will assume that the minimax design  $\hat{\epsilon}$  does not satisfy (1), that is,  $L[D(\hat{\epsilon})] > \min_i L[D(\epsilon)]$ . Then a design  $\epsilon$  can be found such that

$$(\partial/\partial\alpha)L(\epsilon) < 0 \quad \epsilon = (1-\alpha)\hat{\epsilon} + \alpha\epsilon \quad (2.9.21)$$

As a design  $\epsilon$  it is possible, for example, to choose the design  $\hat{\epsilon}$  minimizing  $L[D(\epsilon)]$ .

Then (2.9.21) will immediately follow from (2.9.2)-(2.9.4) and (cf. Theorem 1.1.12) the inequality

$$\begin{aligned} L[M^{-1}(\hat{\epsilon})] &> L[(1-\alpha)M^{-1}(\hat{\epsilon}) + \alpha M^{-1}(\epsilon)] \\ &\geq L\{[(1-\alpha)M(\hat{\epsilon}) + \alpha M(\epsilon)]^{-1}\} \end{aligned} \quad (2.9.22)$$

From the preceding section of the proof of Theorem 2.9.2, it follows that

$$\max_x \varphi(x, \hat{\epsilon}) \leq \min_{\epsilon} L[D(\epsilon)]. \quad (2.9.23)$$

From this and (2.9.11) we obtain that

$$\begin{aligned} \partial L[D(\hat{\epsilon})]/\partial \alpha |_{\alpha=0} &= L[D(\hat{\epsilon})] - L[D(\hat{\epsilon}) M(\epsilon) D(\hat{\epsilon})] \\ &= L[D(\hat{\epsilon})] - \sum_{i=1}^n p_i \lambda(x_i) L[D(\hat{\epsilon}) f(x_i) f'(x_i) D(\hat{\epsilon})] \\ &\geq L[D(\hat{\epsilon})] - \max_x \varphi(x, \hat{\epsilon}) \geq 0, \end{aligned} \quad (2.9.24)$$

where  $x_1, x_2, \dots, x_n$  is the spectrum of the design  $\epsilon$ . The contradiction obtained [cf. (2.9.21) and (2.9.24)] proves our assertion.

(3) We assume that the design  $\hat{\epsilon}$  satisfying condition (3) is not linear-optimal. Then a design  $\epsilon$  can be found for which inequality (2.9.21) will hold. But by (2.9.19) and the definition of the design  $\hat{\epsilon}$ ,

$$\begin{aligned} \partial L[D(\hat{\epsilon})]/\partial \alpha |_{\alpha=0} &= L[D(\hat{\epsilon})] - \sum_{i=1}^n p_i \varphi(x_i, \hat{\epsilon}) \\ &\geq L[D(\hat{\epsilon})] - L[D(\hat{\epsilon})] \sum_{i=1}^n p_i = 0. \end{aligned} \quad (2.9.25)$$

The obtained contradiction [cf. (2.9.25) and (2.9.21)] proves the assertion that if the design  $\hat{\epsilon}$  satisfies (3) then it is linear-optimal and it follows minimax.

(4) The concluding part of the theorem follows immediately from the convexity of the function  $L[M^{-1}(\epsilon)]$ . The theorem is proved.

In the conditions of Theorem 2.9.2, it was assumed that (2.9.4) held. If condition (2.9.4) is replaced by a strict inequality  $L(A) > 0$  for any positive-semidefinite matrix  $A$ , then the result of Theorem 2.9.2 can be somewhat strengthened.

**Theorem 2.9.2a.** *The following assertions:*

- (1) *The design  $\hat{\epsilon}$  minimizes  $L[D(\epsilon)]$ ,*
- (2) *The design  $\hat{\epsilon}$  minimizes  $\max_x \varphi(x, \epsilon)$ ,*
- (3)  $\max_x \varphi(x, \hat{\epsilon}) = L[D(\hat{\epsilon})]$

are equivalent. The information matrix of the designs satisfying (1)–(3) coincide among themselves. Any linear combination of designs satisfying (1)–(3) also satisfies (1)–(3).

The uniqueness of the information matrix of linear-optimal designs follows from the strict convexity of the function  $L[M^{-1}(\epsilon)]$  when inequality (2.9.7) is satisfied.

We note that under the weaker condition (2.9.4) it is possible that optimal designs exist having distinct information matrices.

Theorem 2.9.2 is the basis of our methods of constructing linear-optimal designs. Relying on the results of this theorem, it is possible to verify easily optimality of a given design or not. In this case, it is advantageous to use the following properties of linear-optimal designs.

**Corollary 1** At the point  $\hat{x}_i$  ( $i = 1, 2, \dots, n$ ) of the spectrum of the linear optimal design  $\hat{\epsilon}$

$$\varphi(x_i, \epsilon) = L[D(\hat{\epsilon})] \quad (2.9.26)$$

We assume the contrary. Let  $x_j$  be a point of the spectrum, at which

$$\varphi(x_j, \hat{\epsilon}) < L[D(\hat{\epsilon})]$$

Then, transferring a sufficiently small part  $\alpha$  of the allocation to the point  $x_j$ , where  $\varphi(x, \hat{\epsilon}) = L[D(\hat{\epsilon})]$  (such a point can be found in view of the validity of Theorem 2.9.2), we obtain

$$L[D(\hat{\epsilon}')] = L[D(\hat{\epsilon})] - [\varphi(x, \hat{\epsilon}) - \varphi(x_j, \hat{\epsilon})] \alpha < L[D(\hat{\epsilon})] = \min_{\epsilon} L[D(\epsilon)]$$

The contradiction obtained proves our assertion.

We note that the quantities  $\varphi(x, \epsilon)$  taking part in the preceding considerations are always nonnegative. Indeed,

$$\varphi(x, \epsilon) = \lambda(x)L[D(\epsilon)f(x)f'(x)D(\epsilon)] = \lambda(x)L[AA']$$

where  $A = D(\epsilon)f(x)$ . But the matrix of the form  $AA'$  is positive semidefinite (cf. Theorem 1.1.9), and it follows, in view of the definition of the functional  $L$ , that

$$\varphi(x, \epsilon) = \lambda(x)L[AA'] \geq 0 \quad (2.9.27)$$

**EXAMPLE** We consider the case when the utmost accuracy is required

of one parameter. Without loss of generality, we will denote it by  $\theta_1$ . The functional  $L$  acts in this case on the matrix  $D(\hat{\theta})$  as

$$l'D(\hat{\theta})l, \quad (2.9.28)$$

where

$$l' = \|1, 0, \dots, 0\|, \quad \theta_1 = l'\theta.$$

It is easy to verify that

$$l'(A + B)l = l'Al + l'Bl,$$

$$l'kAl = k l'Al,$$

$$l'Al \geq 0$$

for any positive-semidefinite matrix  $A$ . The function  $\varphi(x, \epsilon)$  has the form

$$\varphi(x, \epsilon) = \lambda(x) l'D(\epsilon) f(x) f'(x) D(\epsilon) l = \lambda(x) \left[ \sum_{\alpha=1}^m D_{1\alpha} f_\alpha(x) \right]^2.$$

In order to find the optimal design, it is sufficient to construct a polynomial  $\lambda^{1/2}(x) \sum_{\alpha=1}^m D_{1\alpha}(\epsilon) f_\alpha(x)$ , which is between  $\pm D_{11}^{1/2}(\epsilon)$ . Sometimes it is more useful instead of the function  $\varphi(x, \epsilon)$  to use the function  $\tilde{\varphi}(x, \epsilon) = \varphi(x, \epsilon) D_{11}^{-1}(\epsilon)$  and correspondingly to contain in the interval  $\pm 1_m$  the polynomial  $\lambda^{1/2}(x) \sum_{\alpha=1}^m D_{1\alpha}(\epsilon) D_{11}^{-1/2}(\epsilon) f_\alpha(x)$  or the polynomial  $\sum_{\alpha=1}^m D_{1\alpha}(\epsilon) D_{11}^{-1/2}(\epsilon) f_\alpha(x)$  in the interval  $\pm \lambda^{-1/2}(x)$ . It is not difficult to obtain (as we might expect) that the condition  $\max_x \tilde{\varphi}(x, \epsilon) \leq 1$  coincides with the condition  $\max_x \lambda(x) d(x, 1, \epsilon) \leq 1$  (cf. Section 2.7). Indeed,

$$\begin{aligned} \lambda(x) d(x, 1, \epsilon) &= \lambda(x) f'(x) \begin{vmatrix} D_{11}(\epsilon) & D_{1(m-1)}(\epsilon) \\ D_{(m-1)1}(\epsilon) & D_{(m-1)(m-1)}(\epsilon) - M_{(m-1)(m-1)}^{-1}(\epsilon) \end{vmatrix} f(x) \\ &= \lambda(x) f'(x) D(\epsilon) \begin{vmatrix} D_{11}^{-1}(\epsilon) & 0 \\ 0 & 0 \end{vmatrix} D(\epsilon) f(x) = \tilde{\varphi}(x, \epsilon). \end{aligned} \quad (2.9.29)$$

In (2.9.29) we made use of the fact that

$$D(\epsilon) \begin{vmatrix} D_{11}^{-1}(\epsilon) & 0 \\ 0 & 0 \end{vmatrix} D(\epsilon) = \begin{vmatrix} D_{11}(\epsilon) & D_{1(m-1)}(\epsilon) \\ D_{(m-1)1}(\epsilon) & D_{(m-1)1}(\epsilon) D_{11}^{-1}(\epsilon) D_{1(m-1)}'(\epsilon) \end{vmatrix}$$

and the Frobenius formula (cf. Section 1.1).

Let

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 \lambda(x) = (1 - |x|)^2, \quad -1 \leq x \leq 1$$

The graph of the function  $\lambda(x)$  is presented in Fig. 11A.

We will try to find a nondegenerate design minimizing the dispersion  $\theta_3$ . It is evident that the function  $P(x) = \sum_{\alpha=1}^3 D_{3\alpha}(\epsilon) D_{33}^{-1/2}(\epsilon) x^{\alpha-1}$  is a polynomial of the second degree. We note that this polynomial's coefficient of  $x^2$  cannot be zero, since this would indicate that  $D_{33}(\epsilon) = 0$ .

At the points of the spectrum of the optimal design the following equality must be satisfied

$$\left| \sum_{\alpha=1}^3 D_{3\alpha}(\epsilon) D_{33}^{-1/2}(\epsilon) x^{\alpha-1} \right| = (1 - |x|)^{-1}$$

Since a nondegenerate design is sought, there must be at least three points. From consideration of the symmetry (cf. Fig. 11B) it follows that  $P(x) = -1 + bx^2, b \geq 0$ .

At points of contact, the following equations must be satisfied

$$\begin{aligned} -1 + bx^2 &= (1 - x)^{-1}, & 2bx &= (1 - x)^{-2}, & 1 \geq x \geq 0, \\ -1 + bx^2 &= (1 + x)^{-1} & -2bx &= (1 + x)^{-2}, & 0 \geq x \geq -1 \end{aligned}$$

It is not difficult to verify that the unique solution of this system is (for  $b > 0$ )

$$x_1 = (7 - (17)^{1/2})/4 \approx 0.72 \quad x_2 = -x_1 \quad b \approx 8.85$$

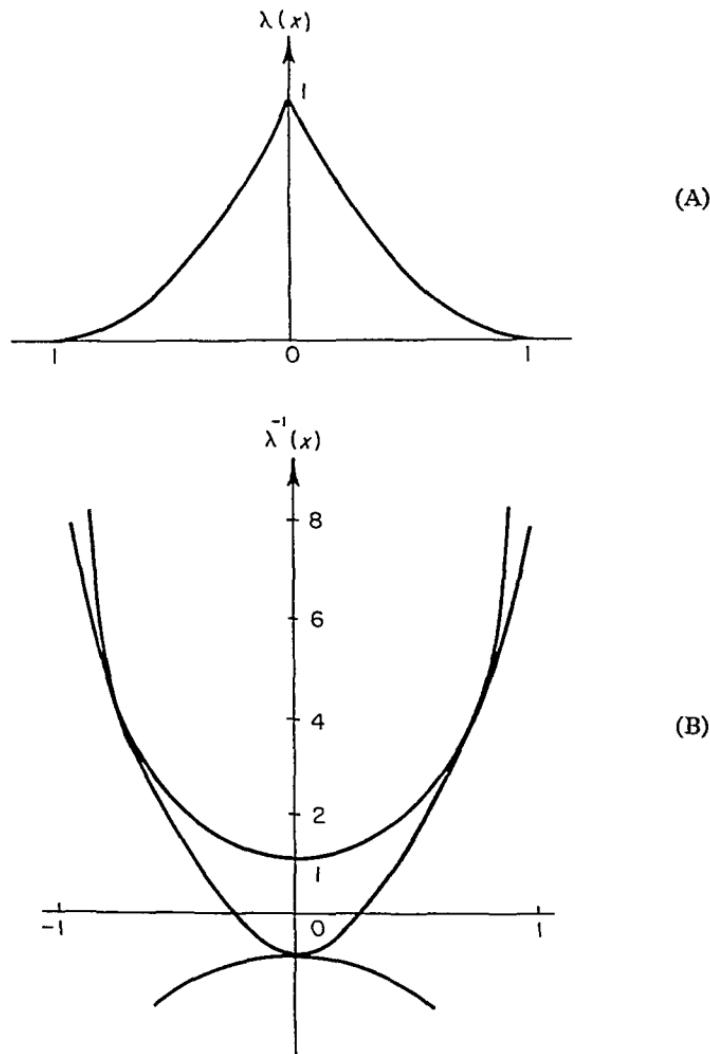
The third point of contact, as is not difficult to see, is the point  $x = 0$ . From this the sought polynomial is  $P(x) = 1 + 8.85x^2$ . Since

$$D_{31}(\epsilon) D_{33}^{-1/2}(\epsilon) + D_{32}(\epsilon) D_{33}^{-1/2}(\epsilon) x + D_{33}^{1/2}(\epsilon) x^2 \equiv 1 + 8.85x^2$$

then  $D_{33}(\epsilon) = 78.3$ ,  $D_{32}(\epsilon) = 0$ , and  $D_{31} = 8.85$ . We will not consider the choice of  $p_i$  ( $i = 1, 2, 3$ ) for the given design here. The resolution of such questions has been given in Section 2.6. We note that for the uniform design

$$\epsilon = \left\{ \begin{array}{l} x_1 = -\frac{1}{2}, \quad x_2 = 0, \quad x_3 = \frac{1}{2} \\ p_1 = p_2 = p_3 = \frac{1}{3} \end{array} \right\}$$

the dispersion of the estimate of the third parameter  $D_{33} = 144$ .



**Fig. 11A.** Effectiveness function  $\lambda(x) = (1 - |x|)^2$ .

**Fig. 11B.** An example of the geometric construction of a design minimizing the variance of one of the parameters.

**EXAMPLE.** We show that there exists linear-optimal designs, which for one and the same regression problem have distinct information matrices.

Let  $\eta(x, \theta) = \theta_1 + \theta_2 x$ ,  $\lambda(x) \equiv 1$ , and  $-1 \leq x \leq 1$ . We consider two designs:

$$\epsilon_1 = \begin{cases} x_1 = 0 \\ p_1 = 1 \end{cases} \quad \text{and} \quad \epsilon_2 = \begin{cases} x_1 = -\gamma, & x_2 = \gamma \\ p_1 = \frac{1}{2}, & p_2 = \frac{1}{2} \end{cases}, \quad 0 < \gamma \leq 1.$$

For the design  $\epsilon_1$  we have

$$M(\epsilon_1) = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} \quad \text{and} \quad D(\epsilon_1, \xi) = \begin{vmatrix} (1+\xi)^{-1} & 0 \\ 0 & \xi^{-1} \end{vmatrix}$$

We verify that the degenerate design  $\epsilon_1$  minimizes  $D(\theta_1) = D_{11}(\epsilon)$ . For this it is sufficient to verify that

$$\max_x \varphi(x, \epsilon_1) = 1$$

We compute  $\varphi(x, \epsilon_1)$

$$\begin{aligned} \varphi(x, \epsilon_1) &= \lim_{\xi \rightarrow 0} \|1 - 0\| \begin{vmatrix} (1+\xi)^{-1} & 0 \\ 0 & \xi^{-1} \end{vmatrix} \begin{vmatrix} 1 & x \\ x & 1-x \end{vmatrix} \begin{vmatrix} (1+\xi)^{-1} & 0 \\ 0 & \xi^{-1} \end{vmatrix} \begin{vmatrix} 1 \\ 0 \end{vmatrix} \\ &= \lim_{\xi \rightarrow 0} (1+\xi)^{-2} = 1 \end{aligned}$$

Since  $\max_x 1 = 1$ , the design  $\epsilon_1$  is optimal.

For the design  $\epsilon_2$  we have

$$M(\epsilon_2) = \begin{vmatrix} 1 & 0 \\ 0 & \gamma^2 \end{vmatrix}, \quad D(\epsilon_2) = \begin{vmatrix} 1 & 0 \\ 0 & \gamma^{-2} \end{vmatrix}.$$

and

$$\begin{aligned} \max_x \varphi(x, \epsilon_2) &= \max_x \|1 - 0\| \begin{vmatrix} 1 & 0 \\ 0 & \gamma^{-2} \end{vmatrix} \begin{vmatrix} 1 \\ x \end{vmatrix} \begin{vmatrix} 1 & x \\ 1-x & 1 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & \gamma^{-2} \end{vmatrix} \begin{vmatrix} 1 \\ 0 \end{vmatrix} \\ &= \max_x 1 = 1 \end{aligned}$$

In this manner both designs  $\epsilon_1$  and  $\epsilon_2$  are optimal, but their information matrices are distinct. We note that in view of Theorem 2.9.2 a linear combination of the design  $\epsilon_1$  and  $\epsilon_2$  also gives an optimal design.

## 2.10. Iterative Methods for Constructing Linear-Optimal Designs

I. Analytic construction of linear-optimal designs is possible only in the simplest cases. These will be considered under the investigation of concrete linear criteria of comparing designs. In this section we will present numerical methods for constructing optimal designs which will rely basically on the results of Theorem 2.9.2.

Let there be some nondegenerate design  $\epsilon_0$ . We consider the design  $\epsilon_1 = (1-\alpha)\epsilon_0 + \alpha\varphi(x)$ . By Lemma 2.9.2, for small  $\alpha$ ,

$$\begin{aligned} L[D(\epsilon_1)] &\simeq L[D(\epsilon_0)] + \alpha \{\partial L[D(\epsilon_1)]/\partial \alpha\} \Big|_{\alpha=0} \\ &= L[D(\epsilon_0)] - \{\varphi(x, \epsilon_0) - L[D(\epsilon_0)]\} \alpha \end{aligned} \quad (2.10.1)$$

If the design  $\epsilon_0$  is not linear-optimal, then by Theorem 2.9.2 there will always be a point  $x$ , where

$$\varphi(x, \epsilon_0) - L[D(\epsilon_0)] > 0. \quad (2.10.2)$$

Moving part of the allocation to the point  $x_0$ , for which the difference  $\varphi(x, \epsilon_0) - L[D(\epsilon_0)]$  attains its maximal value, we obtain a design  $\epsilon_1$  better than the design  $\epsilon_0$ . Repeating such an “improvement” of the design sufficiently often, we eventually reach a time when improvement will be unessential, that is,

$$\frac{L[D(\epsilon_s)] - L[D(\epsilon_{s+1})]}{L[D(\epsilon_{s+1})]} \leq \gamma,$$

where  $\gamma$  is some small positive number. The sequence  $\{L[D(\epsilon_s)]\}$  converges as  $s \rightarrow \infty$  since any decreasing sequence, which is bounded below, converges (in the given case the bound is equal to  $\min_\epsilon L[D(\epsilon)]$ ). Our problem consists of choosing a sequence  $\{\alpha_s\}$  for which the sequence  $\{L[D(\epsilon_s)]\}$  converges to  $\min_\epsilon L[D(\epsilon)]$ .

II. We consider numerical methods for finding linear-optimal designs from a more rigorous mathematical viewpoint.

**Lemma 2.10.1.** *If  $\epsilon$  is any design which is not linear-optimal, then at the points where  $\varphi(x, \epsilon) - L[D(\epsilon)] > 0$  the following inequality holds:*

$$\lambda(x) d(x, \epsilon) - 1 \geq \frac{\varphi(x, \epsilon) - L[D(\epsilon)]}{\varphi(x, \epsilon)} > 0. \quad (2.10.3)$$

*Proof.* From Lemma 2.9.2

$$\begin{aligned} \frac{1}{2}[\partial^2 D(\epsilon)/\partial \alpha^2] \Big|_{\alpha=0} &= D(\epsilon)[\lambda(x)f(x)f'(x) - M(\epsilon)] \\ &\quad \times D(\epsilon)[\lambda(x)f(x)f'(x) - M(\epsilon)] D(\epsilon) \\ &= \lambda^2(x) D(\epsilon) f(x) f'(x) D(\epsilon) f(x) f'(x) D(\epsilon) \\ &\quad - 2\lambda(x) D(\epsilon) f(x) f'(x) D(\epsilon) + D(\epsilon) \\ &= \lambda(x)[\lambda(x)d(x, \epsilon) - 1] D(\epsilon) f(x) f'(x) D(\epsilon) \\ &\quad - [\lambda(x) D(\epsilon) f(x) f'(x) D(\epsilon) - D(\epsilon)] \\ &\geq 0. \end{aligned}$$

Applying the functional  $L$  to both sides of the given matrix inequality we obtain

$$[\lambda(x) d(x, \epsilon) - 1] \varphi(x, \epsilon) - \{\varphi(x, \epsilon) - L[D(\epsilon)]\} \geq 0 \quad (2.10.4)$$

Since  $\varphi(x, \epsilon) - L[D(\epsilon)] > 0$ , then from (2.10.4) it follows that

$$\lambda(x) d(x, \epsilon) - 1 \geq \frac{\varphi(x, \epsilon) - L[D(\epsilon)]}{\varphi(x, \epsilon)} > 0 \quad (2.10.5)$$

The lemma is proved

**Lemma 2.10.2.** *Let the design  $\epsilon_{s+1} = (1 - \alpha) \epsilon_s + \alpha \epsilon(x)$ , then*

$$(1) \quad L[D(\epsilon_{s+1})] = (1 - \alpha)^{-1} \left\{ L[D(\epsilon_s)] - \frac{\alpha \varphi(x, \epsilon_s)}{1 - \alpha + \alpha \lambda(x) d(x, \epsilon_s)} \right\}, \quad (2.10.6)$$

$$(2) \quad L[D(\epsilon_s)] - L[D(\epsilon_{s+1})] > 0,$$

if at the point  $x$ ,

$$\varphi(x, \epsilon_s) - L[D(\epsilon_s)] > 0$$

and

$$0 < \alpha < \frac{\varphi(x, \epsilon_s) - L[D(\epsilon_s)]}{\varphi(x, \epsilon_s)[\lambda(x) d(x, \epsilon_s) - 1]} \quad (2.10.7)$$

*Proof* By Theorem 2.6.1,

$$D(\epsilon_{s+1}) = (1 - \alpha)^{-1} \left[ 1 - \frac{\alpha \lambda(x) D(\epsilon_s) f(v) f(x)}{1 + \alpha [\lambda(x) d(x, \epsilon_s) - 1]} \right] D(\epsilon_s) \quad (2.10.8)$$

Using the definition of the functional  $L$ , from (2.10.8) one may easily obtain, in agreement with (2.10.6), that

$$L[D(\epsilon_{s+1})] = \frac{1}{1 - \alpha} \left\{ L[D(\epsilon_s)] - \frac{\alpha \varphi(x, \epsilon_s)}{1 + \alpha [\lambda(x) d(x, \epsilon_s) - 1]} \right\}$$

From (2.10.6),

$$L[D(\epsilon_s)] - L[D(\epsilon_{s+1})] = \frac{\alpha}{1 - \alpha} \left\{ \frac{\varphi(x, \epsilon_s)}{1 + \alpha [\lambda(x) d(x, \epsilon_s) - 1]} - L[D(\epsilon_s)] \right\}$$

An easy verification shows that

$$L[D(\epsilon_s)] - L[D(\epsilon_{s+1})] > 0$$

if (cf. also Lemma 2.10.1)  $\alpha$  is contained within the limits

$$0 < \alpha < \frac{\varphi(x, \epsilon_s) - L[D(\epsilon_s)]}{\varphi(x, \epsilon_s)[\lambda(x) d(x, \epsilon_s) - 1]}.$$

The lemma is proved.

We now turn our attention to the upper bound [cf. (2.10.7)] for the values of the coefficient  $\alpha$ , for which the design  $\epsilon_{s+1}$  is preferred to the design  $\epsilon_s$ , and which is always no larger than 1, if  $\varphi(x, \epsilon_s) - L[D(\epsilon_s)] > 0$ . Indeed, by Lemma 2.10.1,

$$\alpha = \frac{\varphi(x, \epsilon_s) - L[D(\epsilon_s)]}{\varphi(x, \epsilon_s)[\lambda(x) d(x, \epsilon_s) - 1]} \leq \frac{\varphi(x, \epsilon_s)}{\varphi(x, \epsilon_s)} = 1. \quad (2.10.9)$$

We consider the following iterative procedure.

1. Some nondegenerate design  $\epsilon_s$  with dispersion matrix  $D(\epsilon_s)$  is available.
2. The point  $x_s$  is sought corresponding to

$$\max_x \{\varphi(x, \epsilon_s) - L[D(\epsilon_s)]\}.$$

3. The design  $\epsilon_{s+1} = (1 - \alpha_s) \epsilon_s + \alpha_s \epsilon(x_s)$  is constructed, where

$$\alpha_s = \frac{\varphi(x_s, \epsilon_s) - L[D(\epsilon_s)]}{\gamma \varphi(x_s, \epsilon_s)[\lambda(x_s) d(x_s, \epsilon_s) - 1]}, \quad (2.10.10)$$

and  $\gamma$  is some constant greater than 1.

4.  $D(\epsilon_{s+1})$  and  $L[D(\epsilon_{s+1})]$  are computed.

Furthermore, all operations 1–3 are repeated with the design  $\epsilon_{s+1}$ , and so on

**Theorem 2.10.1.** *The sequence  $\{L[D(\epsilon_s)]\}$  converges:*

$$\lim_{s \rightarrow \infty} L[D(\epsilon_s)] = L[D(\tilde{\epsilon})].$$

If the design  $\tilde{\epsilon}$  is not degenerate ( $|D(\tilde{\epsilon})| \neq 0$ ), then

$$L[D(\tilde{\epsilon})] = \min_{\epsilon} L[D(\epsilon)].$$

*Proof* If  $\alpha_s$  is chosen in agreement with (2.10.10), then by Lemma 2.10.2

$$L[D(\epsilon_0)] \geq L[D(\epsilon_1)] \geq \dots \geq L[D(\epsilon_s)] \geq \dots \geq \min_{\epsilon} L[D(\epsilon)]$$

But each monotone-decreasing sequence which is bounded below converges. We will show that for a nondegenerate design  $\bar{\epsilon}$

$$L[D(\epsilon)] = \min_{\epsilon} L[D(\epsilon)]$$

If this is not so, then the design  $\bar{\epsilon}$  is distinct from the linear-optimal one, and by Theorem 2.9.2,

$$\max_x \{q(x, \epsilon) - L[D(\epsilon)]\} = A > 0$$

Since the design  $\epsilon$  is nondegenerate, then  $\max_x \lambda(x) d(x, \epsilon) = c < \infty$ . Therefore

$$\lim_{s \rightarrow \infty} \alpha_s = A / [\alpha q(x, \epsilon)(c - 1)] \geq v > 0$$

But, on the other hand, for  $\max_x \lambda(x) d(x, \epsilon) = c < \infty$ , from formula (2.10.6) and the necessary condition for convergence

$$\lim_{s \rightarrow \infty} \{L[D(\epsilon_s)] - L[D(\epsilon_{s+1})]\} = 0$$

It follows that  $\lim_{s \rightarrow \infty} \alpha_s = 0$ . The contradiction obtained proves our assertion. The theorem is proved.

*Remark 1* From the proof of the theorem, it is not difficult to obtain that the iterative procedure presented will converge only to a linear optimal design, if for the initial design  $\epsilon_0$  the condition

$$L[D(\epsilon_0)] < \min_{\epsilon} I[D(\epsilon)] \quad (2.10.11)$$

is satisfied, where  $\mathcal{D}$  is the set of nondegenerate designs for which

$$L[D(\epsilon)] > \min_{\epsilon} I[D(\epsilon)]$$

III. It is possible to show that the iterative procedure 1–4 converges for other choices of the sequence  $\alpha_s$ . In the same way as for the construction of  $D$  optimal designs,  $\alpha_s$  must be chosen sufficiently small in order that the sequence  $\{L[D(\epsilon_s)]\}$  decrease with increasing  $s$  and  $\alpha_s$  must not decrease too rapidly in order that the sequence  $\{L[D(\epsilon_s)]\}$

does not converge to a limit for which  $\varphi(\epsilon_s) - L[D(\epsilon_s)] > 0$ . In order to satisfy the last condition, it is sufficient to require that

$$\sum_{s=0}^{\infty} \alpha_s = \infty, \quad \lim_{s \rightarrow \infty} \alpha_s = 0. \quad (2.10.12)$$

The simplest sequence satisfying (2.10.12) is  $\alpha_s \sim s^{-1}$ .

As is shown in practice the more satisfactory results, in the sense of the rapidity of convergence, of the iterative procedure are obtained for the sequence  $\alpha_s$ , chosen in agreement with (2.10.10), and for  $\alpha_s$ , decreasing in  $\gamma > 1$  as soon as

$$L[D(\epsilon_{s+1})] \geq L[D(\epsilon_s)].$$

In these cases the rapidity of converges depends on the value  $\gamma$ .

The exits from the iterative process 1–4 can take place according to one of the rules:

$$\alpha_s \leq \delta_1, \quad \frac{L[D(\epsilon_s)] - L[D(\epsilon_{s+1})]}{L[D(\epsilon_{s+1})]} \leq \delta_2, \quad \max_x \{\varphi(x, \epsilon_s) - L[D(\epsilon_s)]\} \leq \delta_3,$$

where  $\delta_i$  ( $i = 1, 2, 3$ ) are some accuracy of approximations given beforehand.

If the limiting design  $\tilde{\epsilon}$  is degenerate, then it is recommended that the inequality

$$\varphi(x, \tilde{\epsilon}) - L[D(\tilde{\epsilon})] \leq 0, \quad x \in X \quad (2.10.13)$$

be verified (cf. also the corollary to Theorem 2.9.2). In those cases when (2.10.13) is violated, the iterative procedure must be repeated with different initial approximations.

For concrete realizations of the investigated iterative methods on computing machines, the same remarks which were directed to iterative procedures for the construction of  $D$ -optimal designs are to be recommended.

## 2.11. Designs Minimizing $\text{Tr } D(\epsilon)$

I. Experiments minimizing  $\text{Tr } D(\epsilon)$  and corresponding designs, as was noted above, are called  $A$ -optimal. Minimization of  $\text{Tr } D(\epsilon)$  is equivalent to minimization of the mean dispersion of the estimates of the parameters  $D = m^{-1} \sum_{a=1}^m D(\hat{\theta}_a) = m^{-1} \text{Tr } D(\epsilon)$ . From this arises

the name of this criterion of optimality. The requirement of minimum sum of squares of errors of the estimates of the sought parameters  $\text{Tr } D(\epsilon)$  is obvious and natural. Indeed, the obviousness of the given criterion is explained, evidently, by the fact that the experimenter as a rule compares results of various experiments relying on  $\text{Tr } D(\epsilon)$ . Recall that  $[\text{Tr } D(\epsilon)]^{1/2}$  equals half the length of the diagonal of a rectangle circumscribed around the ellipse of dispersion (cf. Fig. 1). From the mathematical point of view this criterion is less useful than the minimax criterion or the criterion of  $D$ -optimality. Thus, for example,  $A$ -optimal designs are not invariant with respect to any nondegenerate linear transformation in the space of estimates of the parameters (compare with Theorem 2.2.4). Methods of constructing  $A$ -optimal designs are somewhat more complicated than methods of constructing  $D$ -optimal designs.

**II** The criterion of  $A$ -optimality, just as the criterion of  $D$ -optimality, depends on the elements of the dispersion matrix  $D(\epsilon)$ . Therefore many properties of  $D$ -optimal designs which follow from the determinant of the dispersion matrix (or the information matrix) remain valid also for  $A$ -optimal designs. In what follows we will study  $A$ -optimal designs using the mathematical apparatus developed in Sections 2.9 and 2.10.

We will show that the functional

$$L[D(\epsilon)] = \text{Tr}_l D(\epsilon) \quad (2.11.1)$$

satisfies conditions (2.9.2)–(2.9.4), where the index  $l$  specifies that the summation takes place only along the diagonal elements of the matrix  $D(\epsilon)$  which correspond to the  $l$  parameters representing the interests of the experimenter ( $l \leq m$ ).

Indeed (cf. Section 1.1)

$$\text{Tr}_l(A + B) = \text{Tr}_l A + \text{Tr}_l B \quad (2.11.2)$$

and

$$\text{Tr}_l kA = k \text{Tr}_l A \quad (2.11.3)$$

It is also obvious that

$$\text{Tr}_l A \geq 0 \quad (2.11.4)$$

for any positive semidefinite matrix  $A$ .

From (2.11.2)–(2.11.4) it follows that for  $A$ -optimal designs Theorems 2.9.1 and 2.9.2 are valid.

Stronger results can be obtained when  $l = m$ . In this case  $\text{Tr } A > 0$  for any positive semidefinite matrix  $A$ , and it follows that Theorem 2.9.2a holds.

The quantity  $\varphi(x, \epsilon)$  introduced in Section 2.9 takes on the form

$$\begin{aligned}\varphi(x, \epsilon) &= \lambda(x) L[D(\epsilon) f(x) f'(x) D(\epsilon)] = \lambda(x) \text{Tr}[D(\epsilon) f(x) f'(x) D(\epsilon)] \\ &= \lambda(x) \text{Tr}[f'(x) D^2(\epsilon) f(x)] = \lambda(x) f'(x) D^2(\epsilon) f(x).\end{aligned}\quad (2.11.5)$$

In (2.11.5) the fact that  $\text{Tr } AB = \text{Tr } BA$  was used. Relying on (2.11.5), it is possible to obtain sufficient conditions for the equivalence of  $D$ - and  $A$ -optimal designs.

**Theorem 2.11.1.** *The design  $\hat{\epsilon}$  is simultaneously  $D$ - and  $A$ -optimal if*

$$kD(\hat{\epsilon}) = D^2(\hat{\epsilon}). \quad (2.11.6)$$

*In this case*

$$\text{Tr } D(\hat{\epsilon}) = km. \quad (2.11.7)$$

*Here  $k$  is some constant and  $m$  is the number of unknown parameters.*

*Proof.* (1) Let  $\hat{\epsilon}$  be a  $D$ -optimal design, then by Theorem 2.2.1 this design minimizes the quantity  $\max_x f'(x) D(\hat{\epsilon}) f(x)$ . Turning attention to (2.11.6) it is not difficult to see that simultaneously the quantity  $\max_x \lambda(x) f'(x) D^2(\hat{\epsilon}) f(x)$  is minimized. But the design minimizing  $\max_x \lambda(x) f'(x) D^2(\hat{\epsilon}) f(x)$  is (cf. Theorem 2.9.2)  $A$ -optimal.

In complete analogy to the previously proved assertion, from  $A$ -optimality follows  $D$ -optimality if (2.11.6) is valid.

(2) According to Theorem 2.2.1 for  $D$ -optimal designs

$$\max_x \lambda(x) f'(x) D(\hat{\epsilon}) f(x) = m. \quad (2.11.8)$$

Using (2.11.6) and Part (3) of Theorem 2.9.2, we obtain

$$\begin{aligned}\text{Tr } D(\hat{\epsilon}) &= \max_x \lambda(x) f'(x) D^2(\hat{\epsilon}) f(x) \\ &= k \max_x \lambda(x) f'(x) D(\hat{\epsilon}) f(x) = km.\end{aligned}\quad (2.11.9)$$

The theorem is proved.

III. The analytic construction of  $A$ -optimal designs is possible only in the simplest cases. Relying on Theorem 2.11.1 it is not difficult to verify that the majority of the constructed designs coincide with the corresponding  $D$ -optimal designs.

For the numerical construction of  $A$ -optimal designs it is useful to use the iterative procedure presented in Section 2.10.

If  $l = m$  (that is, we seek the minimum among the dispersion of all parameters), then the strengthened version of Theorem 2.10.1 is valid.

**Theorem 2.11.2.** *The iterative process 1–4 (cf. Section 2.10) converges, in which case*

$$\lim_{s \rightarrow \infty} \text{Tr } D(\epsilon_s) = \min_{\epsilon} \text{Tr } D(\epsilon)$$

*Proof* By Theorem 2.10.1,

$$\lim_{s \rightarrow \infty} \text{Tr } D(\epsilon_s) = \text{Tr } D(\epsilon) \quad (2.11.10)$$

where the design  $\epsilon$  is either degenerate or optimal.

We will show that if  $l = m$ , then the sequence  $\{\text{Tr } D(\epsilon_s)\}$  cannot converge to a nondegenerate design.

Indeed, if the design  $\epsilon$  is degenerate, then

$$\text{Tr } D(\epsilon) = \infty \quad (2.11.11)$$

But  $\epsilon_0$  is chosen nondegenerate, and it follows that

$$\infty > \text{Tr } D(\epsilon_0) > \text{Tr } D(\epsilon_1) > \dots > \text{Tr } D(\epsilon_s), \quad (2.11.12)$$

or

$$\lim_{s \rightarrow \infty} \text{Tr } D(\epsilon_s) = \text{Tr } D(\epsilon) < \infty$$

The contradiction obtained [cf. (2.11.11) and (2.11.13)] proves the theorem.

The exit from the iterative process for  $l = m$  can occur according to one of the rules

$$\alpha_s < \delta_1,$$

$$\frac{\text{Tr } D(\epsilon_s) - \text{Tr } D(\epsilon_{s+1})}{\text{Tr } D(\epsilon_{s+1})} < \delta_2,$$

$$\max_x \lambda(x) f'(x) D^2(\epsilon_s) f(x) - \text{Tr } D(\epsilon_s) < \delta_3$$

Here  $\delta_i$  ( $i = 1, 2, 3$ ) is an accuracy given beforehand.

In view of the uniqueness of the dispersion matrix corresponding to  $A$ -optimal designs all three rules are closely related to one another and are practically equivalent.

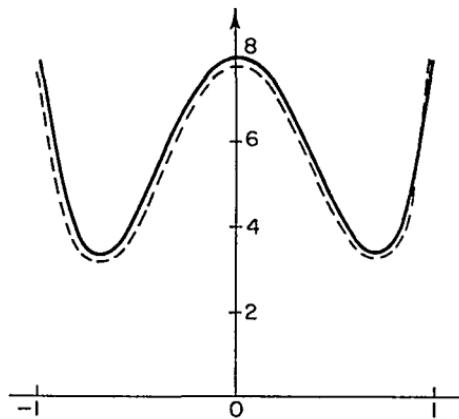
EXAMPLE. Let some quantity  $y$  be measured for  $-1 \leq x \leq 1$  and let

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2, \quad \lambda(x) \equiv 1.$$

It is required to find the design minimizing  $\text{Tr } D(\epsilon)$ . We consider at first analytic methods of construction relying on Theorem 2.9.2.

Corresponding to Corollary 1 of this theorem, for points of the spectrum of the optimal designs  $\epsilon$  the quantity  $\varphi(x, \epsilon) = f'(x) D^2(\epsilon) f(x)$ , where  $f'(x) = \|1, x, x^2\|$ , must be equal to its maximal value which is equal to  $\text{Tr } D(\epsilon)$ . The function  $\varphi(x, \epsilon)$  for the regression problem under consideration is a polynomial of the fourth degree, larger than zero for all  $x$ , and it follows that it has only one extreme point, corresponding to a maximum. From this the spectrum of the design must consist of this point and the two end points  $-1$  and  $1$ . From considerations of symmetry it is obvious that the middle point must equal zero (cf. Fig. 12) and that  $p(-1) = p(1) = p$ .

**Fig. 12.** Behavior of the function  $\varphi(x, \epsilon)$  for the exact and the approximate  $A$ -optimal design.



An easy calculation shows that for the spectrum  $-1, 0, 1$  the sum of the diagonal elements equals (cf. also the example from Section 1.8)

$$\text{Tr } D(\epsilon) = \frac{1}{p(1-2p)}.$$

The minimum of the given expression is obtained for  $p = \frac{1}{2}$  and in this case

$$\varphi(x, \epsilon) = 8 - 20x^2$$

Relying on Theorem 2.9.2 it is not difficult to obtain (cf Fig. 12) that the constructed design

$$\epsilon = \left\{ \begin{array}{l} -1, 0, 1 \\ \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \end{array} \right\}$$

is  $A$  optimal. From the form of the curve  $\varphi(x, \epsilon)$  it follows that this design is unique. For the given problem numerical construction of the optimal design was also carried out. The form of the initial design was chosen to be

$$\epsilon_0 = \left\{ \begin{array}{l} -\frac{1}{2}, 0, \frac{1}{2} \\ \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \end{array} \right\}$$

After 15 iterations we obtained a value of  $\text{Tr } D(\epsilon_{11})$  equal to 8.01. The curve  $\varphi(x, \epsilon_{11})$  is denoted in Fig. 12 by the dashed line.

## 2.12 Designs Minimizing the Mean Dispersion of the Estimate of the Response Surface over the Domain of Values

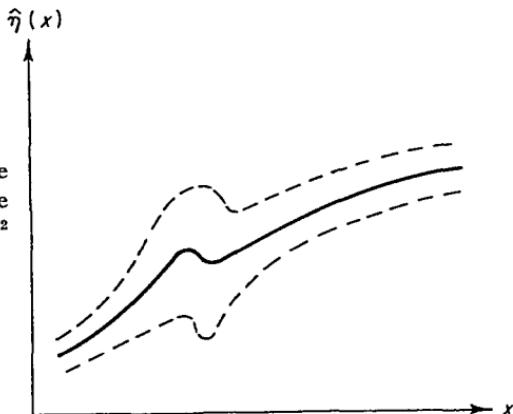
I. If the experimenter is interested in the general regularity of the dependence of the quantities under study on the control variables that is he can sacrifice accuracy of description in small regions for the sake of a good description in the entire region (cf Fig. 13) then it is wise to require the minimization of the quantity

$$\left( \int_Z d\tau \right)^{-1} E \left\{ \int_Z [\eta(x, \theta) - \eta(x, \theta_0)]^2 dx \right\} = \left( \int_Z d\tau \right)^{-1} \int_Z d(x, \epsilon) dx \quad (2.12.1)$$

where  $Z$  generally speaking does not coincide with the region  $X$  of possible observations. The design of the experiment consists in finding a design  $\epsilon$  giving

$$\min \int_Z d(x, \epsilon) dx \quad \min Q(\epsilon) \quad (2.12.2)$$

In the future the design  $\epsilon$  will be called  $Q$  optimal.



**Fig. 13.** The solid line is a plot of the estimate of the curve  $\eta(x)$ ; the distance between the dashed lines equals  $2[d(x)]^{1/2}$  (standard deviation).

II. The functional  $Q(\epsilon)$  satisfies conditions (2.9.2) and (2.9.3). If the functions  $f'(x) = \|f_1(x), f_2(x), \dots, f_m(x)\|$  are linearly independent in the region  $Z$ , then the functional  $Q(\epsilon)$  also satisfies condition (2.9.7) and it follows that Theorem 2.9.2a holds.

For utilization of the results of Theorem 2.9.2 for verifying the  $Q$ -optimality of designs, it is useful to use the following expressions for the quantity  $Q[D(\epsilon)f(x)f'(x)D(\epsilon)]$ :

$$1. \quad Q[D(\epsilon)f(x)f'(x)D(\epsilon)] = \int_Z d^2(x, \tilde{x}, \epsilon) d\tilde{x}, \quad (2.12.3)$$

where  $d(x, \tilde{x}, \epsilon) = f'(x) D(\epsilon) f(\tilde{x})$  is the covariance of the estimator of the response surface at the points  $x$  and  $\tilde{x}$ .

$$2. \quad Q[D(\epsilon)f(x)f'(x)D(\epsilon)] = f'(x) D(\epsilon) \bar{M} D(\epsilon) f(x), \quad (2.12.4)$$

where  $\bar{M} = \int_Z f(x) f'(x) dx$ . Formulas (2.13.3) and (2.12.4) are easy to obtain using the definition of the operator  $Q$  and considering that  $f'(x) D(\epsilon) f(\tilde{x}) = f'(\tilde{x}) D(\epsilon) f(x)$ . Relying on (2.12.4) it is possible to obtain a sufficient condition for the equivalence of  $Q$ - and  $A$ -optimal designs and  $Q$ - and  $D$ -optimal designs (for  $l = m$ ).

**Theorem 2.12.1.** *The design  $\epsilon$  is simultaneously*

(1)  *$Q$ - and  $A$ -optimal if the functions  $f(x)$  are orthogonal in the region  $Z$ .*

(2)  $Q$ - and  $D$ -optimal if  $D(\epsilon) \bar{M} D(\epsilon) = kD(\epsilon)$ , in this case  $\max_x Q(\epsilon) = km$ , where  $k$  is some constant

*Proof* (1) If the functions  $f'(x) = \|f_1(x), f_2(x), \dots, f_m(x)\|$  are orthonormal, then

$$\int_Z f(x) f'(x) dx = I_m, \quad (2.12.5)$$

and

$$D(\epsilon) \bar{M} D(\epsilon) = D(\epsilon) I_m D(\epsilon) = D^2(\epsilon) \quad (2.12.6)$$

From (2.12.6) and Theorem 2.9.2 it immediately follows that Part (1) is valid. This result can be obtained relying immediately on the definition of  $A$ - and  $Q$ -optimal designs. Indeed,

$$\begin{aligned} Q(\epsilon) &= \int_Z f(x) D(\epsilon) f(x) dx = \int_Z \text{Tr}[D(\epsilon) f(x) f'(x)] dx \\ &= \int_Z \text{Tr}[D(\epsilon) I_m] d\tau = \text{Tr } D(\epsilon) \end{aligned}$$

which is what was required to be shown.

(2) The proof of the second part of the theorem is completely analogous to that of Theorem 2.11.1

Theorem 2.12.1 permits a comparatively simple division of  $D$ -optimal designs into designs which are simultaneously also  $Q$ -optimal.

III. Since  $\int_Z f'(x) D(\epsilon) f(x) dx < \infty$  only if  $D(\epsilon)$  is not degenerate, then, in the same way as for  $A$  optimal designs, the following theorem holds

**Theorem 2.12.2.** *The iterative process 1-4 (cf. Section 2.10) converges, in which case*

$$\lim_{i \rightarrow \infty} \int_Z d(x, \epsilon_i) dx = \min_i \int_Z d(x, \epsilon) d\tau$$

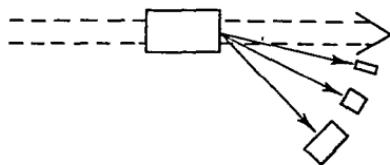
The proof of the theorem is completely analogous to the proof of Theorem 2.11.2

### 2.13. Extrapolation at a Point

I. In many experimental problems, it is necessary to know the dependence of the quantity under study on the control variables in those regions where observations are impossible to obtain or difficult from the practical point of view [small effectiveness  $\lambda(x)$ ]. In these cases there is nothing else to do but to study the behavior of the quantities of interest to us in a complementary region and afterwards extrapolate the obtained dependence into the original region.

EXAMPLE 1. As one of numerous examples one may consider an experiment in the scattering of elementary particles. It is well known that for many theoretical investigations in the physics of elementary particles it is necessary to know the differential cross section for the zero scatter. At the same time, it is impossible to take measurements of the differential section for an angle of scatter  $x$  equal to zero, since the unscattered particles of the primary bundle fall into the counter. For small angles  $x$ , in order for the counters to be separated from a background of unscattered particles, their dimension must be decreased. This causes a small number of scattered particles to fall into the counter, which, in turn, causes an increase of the error  $b(y)$  in determining the differential section of scatter  $y$  (cf. Fig. 14).

**Fig. 14.** Diagram of a simple experiment in the scattering of elementary particles. The dotted lines denote the primary path.



In [5] it is shown that

$$b^2(y_i) = \eta_0^2(x_i)/n_i,$$

where  $n_i$  is the number of particles falling in the counter,  $\eta_0(x_i)$  is the mean value of  $y$  in the region covered by the counter. Assuming that  $\eta_0(x)$  changes slowly in regions close to  $x = 0$ , it is easy to see that

$$b^2(y_i) \sim n_i^{-1} \sim x_i^{-2}.$$

In other words, close to zero, the efficiency of the experiment in measuring the differential section is proportional to  $x^2$ . Therefore, on the one hand, for increasing accuracy in determination of  $\eta_0(0)$  it is natural to strive to place the observations as close as possible to

$x = 0$ . On the other hand, this is not advantageous, since the efficiency  $\lambda(x) \sim x^2$ , and equals zero for  $x = 0$ . Therefore it is necessary to find a compromise solution. The solution of such a problem will be investigated presently.

II. Let it be necessary to extrapolate the experimental dependence to a given point  $x_0$ . It is obvious that it is necessary to aim at making the dispersion  $d(x_0, \epsilon)$  of the predicted value  $\hat{\eta}(x_0, \theta)$  as small as possible. Therefore we will consider the design  $\epsilon$  optimal if it minimizes the quantity  $d(x_0, \epsilon)$ . Since the dispersion  $d(x_0, \epsilon)$ , where  $\epsilon$  is a given nondegenerate design, equals  $f'(x_0) D(\epsilon) f(x_0)$  (cf. Section 1.3), it is obvious that

$$\begin{aligned} f(x_0)(A + B)f(x_0) &= f(x_0) Af(x_0) + f(x_0) Bf(x_0), \\ f(x_0) kAf(x_0) &- kf(x_0) Af(x_0), \end{aligned} \quad (2.13.1)$$

and

$$f(x_0) Af(x_0) \geq 0 \quad (2.13.2)$$

for any positive-definite matrix  $A$ . From (2.13.1) and (2.13.2) it follows that for designs minimizing the dispersion at the given point, we may apply the results obtained in Sections 2.9 and 2.10. In what follows Theorem 2.9.2 will be particularly important. The quantity  $L[D(\epsilon)f(x)f'(x)D(\epsilon)]$  in the formulation of this theorem will have the form

$$L[D(\epsilon)f(x)f'(x)D(\epsilon)] = d^2(x, x_0, \epsilon), \quad (2.13.3)$$

where  $d(x, x_0, \epsilon) = f(x)D(\epsilon)f(x_0) = f'(x_0)D(\epsilon)f(x)$  is the covariance of the estimators  $\hat{\eta}(x_0, \theta)$  and  $\hat{\eta}(x, \theta)$ .

III. If some bounds, in the region  $X$  of possible observations, are placed on the functions  $f'(x) = \|f_1(x_1), f_2(x), \dots, f_m(x)\|$  and the efficiency  $\lambda(x)$ , then the class of designs  $\epsilon$ , among which it is necessary to search for the optimal design, can be reduced.

**Theorem 2.13.1.** *If the optimal design  $\epsilon$  has the number of points  $n = m$ , then the allocation at the  $i$ th point of the design must be equal to*

$$p_i = \frac{|L_i(x_0)| \lambda^{-1/2}(x_i)}{\sum_{j=1}^m |L_j(x_0)| \lambda^{-1/2}(x_j)} \quad (2.13.4)$$

*Proof.* Corresponding to Corollary 1 of Theorem 2.9.2, at the points of the design  $\hat{\epsilon}$

$$\lambda(x_i) d^2(x_i, x_0, \epsilon) = d(x_0, \epsilon) = \text{const}, \quad (i = 1, 2, \dots, m). \quad (2.13.5)$$

Passing to the Lagrange interpolation polynomial

$$L_i(x) = \frac{|F(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_m)|}{|F(x_1, \dots, x_m)|}, \quad (2.13.6)$$

where

$$F(x_1, \dots, x_m) = \|f(x_1), \dots, f(x_m)\|,$$

we may easily verify (cf. Section 1.3) that

$$\begin{aligned} \hat{\eta}(x, \theta) &= \sum_{i=1}^m y_i L_i(x), \\ d(x, \epsilon) &= \sum_{i=1}^m L_i^2(x)/p_i \lambda(x_i), \\ d(x, x_0, \epsilon) &= \sum_{i=1}^m L_i(x) L_i(x_0)/p_i \lambda(x_i). \end{aligned} \quad (2.13.7)$$

Since  $L_i(x_j) = \delta_{ij}$  ( $\delta_{ii} = 1$ ,  $\delta_{ij} = 0$ ,  $i \neq j$ ), then

$$\lambda(x_i) d^2(x_i, x_0, \epsilon) = L_i^2(x_0)/p_i^2 \lambda(x_i) = \text{const}. \quad (2.13.8)$$

From this and the normalization condition  $\sum_{i=1}^m p_i = 1$ ,

$$p_i = \frac{|L_i(x_0)| \lambda^{-1/2}(x_i)}{\sum_{j=1}^m |L_j(x_0)| \lambda^{-1/2}(x_j)},$$

which proves the theorem.

**IV.** More complete results can be obtained (see [32, 33]) when there is only one control variable  $x$ . Let the system of functions

$$\lambda^{1/2}(x) f'(x) = \| \lambda^{1/2}(x) f_1(x), \lambda^{1/2}(x) f_2(x), \dots, \lambda^{1/2}(x) f_m(x) \|$$

be Chebyshev on the interval  $[a, b]$ . Then (cf. [34]) on the interval  $[a, b]$  there exists a polynomial

$$u(x) = \sum_{\alpha=1}^m a_{\alpha}^* \lambda^{1/2}(x) f_{\alpha}(x)$$

such that

$$1 \quad |u(x)| \leq 1, \quad x \in [a, b], \text{ and}$$

2 there are  $m$  points  $a \leq \hat{x}_1 < \hat{x}_2 < \dots < \hat{x}_m \leq b$ , at which  $u(\hat{x}_i) = (-1)^{m-i}$

**Theorem 2.13.2.** If the system of function  $\lambda^{1/2}(x)f(x)$  is Chebyshev on the interval  $[a, b]$ , then the design  $\xi$ , concentrated at the points  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$  with allocation

$$p_i = \frac{|L_i(x_0)| \lambda^{-1/2}(x_i)}{\sum_{j=1}^m |L_j(x_0)| \lambda^{-1/2}(x_j)} \quad (i = 1, 2, \dots, m), \quad (2.13.9)$$

minimizes the dispersion of the estimate at the point  $x_0 \notin [a, b]$

*Proof* We consider the case when  $\lambda(x) \equiv 1$ . The remaining cases are reduced to this by the transformation (cf Section 2.2)  $\varphi(x) = \lambda^{1/2}(x)f(x)$ .

If we assume that the optimal design is concentrated on the  $m$  given points, then by Theorem 2.13.1 the allocation must be chosen as in (2.13.9). In agreement with (2.13.7)

$$\begin{aligned} d(x, x_0, \epsilon) &= \sum_{i=1}^m p_i^{-1} L_i(x) L_i(x_0) \\ &= \sum_{j=1}^m |L_j(x_0)| \sum_{i=1}^m (-1)^{m-i} L_i(x) \end{aligned} \quad (2.13.10)$$

In (2.13.10) we used the interpolation polynomial  $L_i(x)$  and the fact that for a Chebyshev system of functions the determinant  $|F(x_1, \dots, x_m)|$  has one and the same sign [cf (2.3.5)] for any  $a \leq x_1 < x_2 < \dots < x_m \leq b$ .

The polynomial  $u(x)$ , defined above, can be represented in the form

$$u(x) = \sum_{i=1}^m (-1)^{m-i} L_i(x), \quad (2.13.11)$$

where  $L_i(x)$  is the interpolation polynomial constructed at the points  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$ . By definition

$$|u(\hat{x}_i)| = 1 \quad \text{and} \quad |u(x)| \leq 1 \quad (2.13.12)$$

Comparing (2.13.10) and (2.13.11), we see that

$$d(x, x_0, \epsilon) = cu(x).$$

From this and from (2.13.12),

$$d(\hat{x}_i, x_0, \epsilon) = c,$$

and

$$d(x, x_0, \epsilon) \leq c, \quad x \neq \hat{x}.$$

Therefore, for the proof of the theorem it is sufficient to show (cf. Theorem 2.9.2) that

$$d(x_0, \epsilon) = \max_x d^2(x, x_0, \epsilon) = d^2(\hat{x}_i, x_0, \epsilon).$$

From (2.13.7), (2.13.9), and (2.13.10) it follows that

$$d(x_0, \epsilon) = \sum_{i=1}^m L_i^2(x_0) \left[ \sum_{j=1}^m |L_j(x_0)| / |L_i(x_0)| \right] = \left[ \sum_{j=1}^m |L_j(x_0)| \right]^2,$$

and

$$\begin{aligned} d^2(\hat{x}_i, x_0, \epsilon) &= \left[ \sum_{j=1}^m |L_j(x_0)| \sum_{i=1}^m (-1)^{m-i} \delta_{ij} \right]^2 \\ &= \left[ \sum_{j=1}^m |L_j(x_0)| \right]^2. \end{aligned}$$

Therefore

$$d(x_0, \epsilon) = d^2(\hat{x}_i, x_0, \epsilon), \quad i = 1, 2, \dots, m.$$

The theorem is proved.

We note that the points at which it is necessary to take observations for the optimal design does not depend on the point  $x_0$ . Under a translation of  $x_0$  only the distribution of resources changes [cf. (2.13.9)].

**EXAMPLE 2.** Let  $f'(x) = \|1, x, \dots, x^{m-1}\|$ ,  $\lambda(x) \equiv 1$ , and  $[a, b] = [-1, 1]$ . It is well known that the algebraic polynomial deviating least from zero is the Chebyshev polynomial of the first kind (cf. [17]):

$$T_{m-1}(x) = \cos[(m-1) \arccos x].$$

This polynomial on the interval  $[-1, 1]$  has extreme values 1 and  $-1$  and attains these extreme values alternately at the points

$$\hat{x}_i = \cos[(m-i)\pi/(m-1)], \quad (i = 1, 2, \dots, m)$$

Since the points  $\hat{x}_i$ , where  $|T_{m-1}(x)|$  attains its largest value, are  $m$  in number, the design concentrated at them is unique [the design can be concentrated only at points where  $d(x, x_0, \epsilon) = cT_{m-1}(x)$  attains its maxima (cf. Theorem 2.9.2)]. The allocation of observations is determined by formula (2.13.9).

As a concrete example, we choose  $\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2$ . If it is required to know the value of the function at the point  $x = 2$ , then the optimal design has the form

$$\epsilon = \{x_1 = -1, x_2 = 0, x_3 = 1\} \\ p_1 = \frac{1}{3}, \quad p_2 = \frac{1}{3}, \quad p_3 = \frac{1}{3}\}$$

The dispersion of the estimate of the curve  $\eta(x, \theta)$  when  $x = 2$  for the optimal design is equal to  $d(2, \epsilon) = 49$

For the uniform design

$$\epsilon = \{x_1 = -1, x_2 = 0, x_3 = 1\} \\ p_i = \frac{1}{3} \quad (i = 1, 2, 3)\}$$

the dispersion equals  $d(2, \epsilon) = 57$

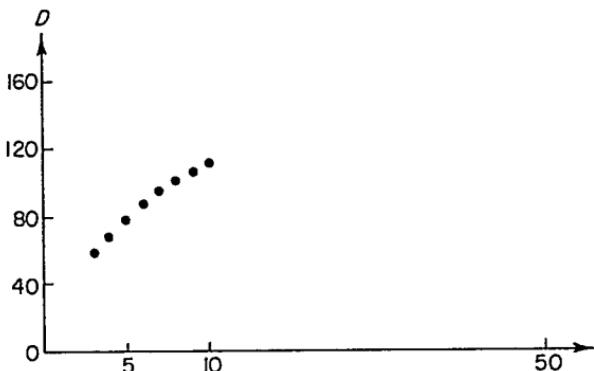
It is interesting that a larger number of points in the spectrum of the uniform design leads to a decrease in accuracy in the determination of the estimate  $\eta(x, \theta)$  (cf. Fig. 15). For  $n = 50$ ,  $d(2, \epsilon) \approx 165$ .

**EXAMPLE 3** In many cases, for a small number of unknown parameters, the problem of extrapolation at a point can be solved geometrically [34]. For this, it is necessary to find the curve

$$u(x) = \lambda^{1/2}(x) \sum_{n=1}^m a_n f_n(x),$$

confined between the values  $\pm 1$ .

Let  $\eta(x, \theta) = \theta_1 + \theta_2 x$  and  $\lambda(x) = x^2$ ,  $0 \leq x \leq 1$ . It is required to find the design minimizing  $d(0, \epsilon)$ . The regression problem formulated is very often met in situations presented in Example 1 of the current part.



**Fig. 15.** The increment of the variance of the predicted value of the response surface at the point  $x = 2$  for increasing number of points in the spectrum of the uniform design.

We now turn our attention to the system of functions

$$\lambda^{1/2}(x)f_1(x) = x, \quad \lambda^{1/2}(x)f_2(x) = x^2, \quad \lambda^{1/2}(x)f_3(x) = x^3,$$

which is not Chebyshev.

In the given concrete case, instead of inscribing

$$u(x) = \lambda^{1/2}(x) \sum_{\alpha=1}^3 a_\alpha f_\alpha(x)$$

between the values  $\pm 1$ , it is advantageous to inscribe  $\tilde{u}(x) = \lambda^{-1/2}(x) u(x)$  between the values  $\pm \lambda^{-1/2}(x)$ . The line with two points of contact with the boundaries  $\pm \lambda^{-1/2}(x)$  is given in Fig. 16. The coordinates of the points of contact are 0.414 and 1.

If we want to obtain the estimate  $\hat{\eta}(x, \theta)$  with smallest dispersion at the point  $x = 0$ , then the distribution of resources must be

$$p_1 = \frac{|L_1(0)|}{|L_1(0)| + |L_2(0)|} = 0.71,$$

and

$$p_2 = \frac{|L_2(0)|}{|L_1(0)| + |L_2(0)|} = 1 - p_1 = 0.29,$$

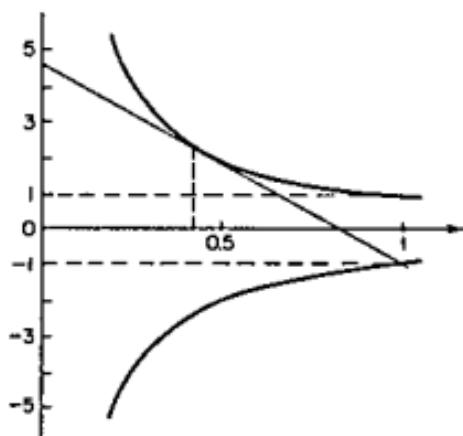


Fig. 16 An example of a geometrical construction of a design minimizing the variance of the regression curve at a given point

where the interpolation polynomial  $L_i(x)$  is

$$L_i(x) = (x - x_j)(x_i - x_j) \quad (i = 1, 2, j \neq i)$$

The dispersion of the curve  $\hat{y}(x, \theta)$  at the point  $x = 0$  is, for the optimal design,

$$d(0, \epsilon) = \sum_{i=1}^2 L_i^2(0) p_i \lambda(x_i) = 14.2$$

If extrapolation of the curve  $\hat{y}(x, \theta)$  is required at other points, for example,  $x_0 < 0$ , then the spectrum of the design remains as before, but the allocation at the points  $x_1$  and  $x_2$  will be determined by the relation

$$p_1/p_2 = |L_1(x_0)|/|L_2(x_0)| = (1 - x_0)(0.414 - x_0)$$

V. Computational methods for constructing designs minimizing the dispersion  $\hat{y}(x, \theta)$  at a given point  $x_0$ , not belonging to  $X$  the space of possible observations are practically identical to the computational methods for the construction of  $A$ - and  $Q$  optimal designs As earlier, the most useful methods appear to be iterative methods for constructing the optimal designs

If the design  $\epsilon_s$  converges to the design  $\epsilon$ , containing  $m$  points, then the allocation of resources is given by formula (2.13.4)

## 2.14. Quadratic Loss

Let the experiment  $\epsilon_1$  be preferred to the experiment  $\epsilon_2$  ( $\epsilon_1 > \epsilon_2$ ), if

$$E[W | \epsilon_1] < E[W | \epsilon_2]. \quad (2.14.1)$$

Here  $E$  is the mathematical expectation operator,

$$W = (\hat{\theta} - \theta)' Q (\hat{\theta} - \theta),$$

and  $Q$  is a positive-semidefinite matrix of dimension  $m \times m$ . The quantity  $E[W | \epsilon]$  is usually called the quadratic loss (cf. [21]) for the experiment  $\epsilon$ .

If the experiment  $\epsilon$  is given, then the quadratic loss is minimized when

$$\hat{\theta} = M^{-1}(\epsilon) Y, \quad (2.14.2)$$

where  $M(\epsilon)$  and  $Y$  are defined by means of formulas (1.3.9) and (1.3.10). It is easy to obtain that  $W = (\hat{\theta} - \theta)' CC' (\hat{\theta} - \theta)$  if it is taken into account that the matrix  $Q$  can be represented in the form  $CC'$  (cf. Theorem 1.1.9). In this case, the problem becomes equivalent to finding the minimum of

$$E(W) = E[(\hat{t} - t)' (\hat{t} - t)] = \text{Tr } D(\hat{t}), \quad (2.14.3)$$

where  $t = C'\theta$ . In view of Corollary 4 to Theorem 1.2.2, the best linear estimate (2.14.2) minimizes (2.14.3) for a given experiment.

We therefore choose a design which minimizes  $E(W)$ , with the estimate constructed according to the rule (2.14.2). Since

$$EW = \text{Tr } D(\hat{t}) = \text{Tr } D(\hat{\theta}) Q, \quad (2.14.4)$$

then a design  $\epsilon_1 > \epsilon_2$ , if

$$\text{Tr } D(\epsilon_1) Q < \text{Tr } D(\epsilon_2) Q. \quad (2.14.5)$$

We verify that the operator acting in (2.14.5) on the matrix  $D(\epsilon)$  satisfies the requirements (2.9.2), (2.9.3), and (2.9.7) or (2.9.4).

Indeed,

$$\text{Tr}(A + B) Q = \text{Tr } A Q + \text{Tr } B Q,$$

$$\text{Tr}(kA) Q = k \text{Tr } A Q$$

for arbitrary matrices  $A$  and  $B$  of dimension  $m \times m$ . Furthermore,  $\text{Tr } A\bar{Q} > 0$  for any positive definite matrix  $A$  and  $\text{Tr } A\bar{Q} \geq 0$  for any positive-semidefinite matrix  $A$ .

From this it follows that the material developed in Sections 2.9 and 2.10 for constructing optimal designs can be applied to the preceding criteria for the comparison of experiments. We note that if  $\bar{Q} = I_m$ , where  $I_m$  is the identity matrix, we arrive at  $A$  optimality which was already considered. If  $Q_{\alpha\alpha} = 1$  and  $Q_{\alpha\beta} = 0$ ,  $\alpha \neq \beta$ , and  $\beta \neq \gamma$ , then we arrive at the case when the experimenter is particularly interested in the parameter  $\theta_\gamma$ . In this case  $D(\theta_\gamma)$  must be minimized. For further enumeration of particular cases, applied in practice, the reader can continue independently.

**EXAMPLE** We now consider the case

$$\bar{Q} = \begin{vmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & 0 \\ 0 & 0 & Q_m \end{vmatrix}$$

The choice of such a matrix  $\bar{Q}$  corresponds to the loss of an inaccurate determination of the estimate of each parameter, and the losses for calculating the relationship between estimates of various parameters are absent. In this case

$$L[D(\epsilon)] = \text{Tr } D(\epsilon)\bar{Q} = \sum_{\alpha=1}^m Q_\alpha D_{\alpha\alpha}(\epsilon) \quad (2.14.6)$$

# 3

## Properties and Methods of Construction for Optimal Discrete Designs

### 3.1. Discrete Designs

In Chapter 2, we investigated the properties of continuous optimal designs. These designs can be considered as approximations of discrete designs  $\epsilon(N)$  (cf. Section 1.10). In this section, we consider the dependence of the accuracy of the approximation of the discrete designs by continuous ones on the possible number of observations. For those cases where this accuracy is not considered satisfactory, we give some recommendations for constructing accurate optimal designs.

By a discrete optimal design, here and in the future, unless explicitly stated otherwise, we will mean the design optimal for a given  $N$ .

I. Let the normalized design  $\epsilon_1$  be better than the normalized design  $\epsilon_2$ , if

$$\tilde{L}[M(\epsilon_1)] > \tilde{L}[M(\epsilon_2)], \quad (3.1.1)$$

where  $M(\epsilon)$  is the information matrix corresponding to the design  $\epsilon$ , and the operator  $\tilde{L}$  satisfies the conditions

$$\tilde{L}(A + B) \geq \tilde{L}(A), \quad \tilde{L}(kA) = k\tilde{L}(A), \quad (3.1.2)$$

where  $A$  and  $B$  are positive-semidefinite matrices. We assume that  $\hat{\epsilon}$  is a continuous normalized design maximizing  $L[M(\epsilon)]$  on the set of continuous designs, and  $\hat{\epsilon}(N)$  is a normalized discrete design maximizing  $L[M[\epsilon(N)]]$  on the set of discrete designs, and let  $M(\hat{\epsilon})$  and  $M[\hat{\epsilon}(N)]$  be their information matrices. Since  $\max_{\epsilon(N)} L[M[\epsilon(N)]]$  is taken over a smaller set than  $\max_{\epsilon} L[M(\epsilon)]$ , it is clear that

$$\max_{\epsilon(N)} L[M[\epsilon(N)]] \leq \max_{\epsilon} L[M(\epsilon)]$$

II. We can find a relation between the two quantities  $\max_{\epsilon(N)} L[M[\epsilon(N)]]$  and  $\max_{\epsilon} L[M(\epsilon)]$  which gives a very useful method for constructing discrete designs [36].

**Theorem 3.1.1.** *If the sum of the number of possible measurements is  $N$ , and  $n$  is the number of points in the spectrum of the design  $\hat{\epsilon}$ , then*

$$[(N - n) N] L[M(\hat{\epsilon})] < L[M[\hat{\epsilon}(N)]] \leq L[M(\epsilon)] \quad (3.1.3)$$

*Proof* Let  $[c]^+$  indicate the smallest integer satisfying the inequality  $[c]^+ \geq c$ . Then by (3.1.2) it is possible to write

$$L \left[ \sum_{i=1}^n [N p_i]^+ f(\hat{x}_i) f'(\hat{x}_i) \right] \geq L \left[ \sum_{i=1}^n N p_i f(\hat{x}_i) f'(\hat{x}_i) \right] = NL[M(\epsilon)] \quad (3.1.4)$$

For  $N_1 > N_2$ ,  $L[N_1 M[\hat{\epsilon}(N_1)]] \geq L[N_2 M[\hat{\epsilon}(N_2)]]$  since for the  $N_1$  measurements the maximum possible amount of information cannot be less than the amount of information extracted from the smaller number of measurements  $N_2$ . From this and (3.1.4)

$$(N + n) L[M[\hat{\epsilon}(N + n)]] > L \left[ \sum_{i=1}^n [N p_i]^+ f(\hat{x}_i) f'(\hat{x}_i) \right] > NL[M(\epsilon)] \quad (3.1.5)$$

or

$$(N + n) L[M(\hat{\epsilon})] \geq (N + n) L[M[\hat{\epsilon}(N + n)]] > NL[M(\epsilon)] \quad (3.1.6)$$

In (3.1.5) and (3.1.6) it was taken into account that

$$\sum_{i=1}^n [N p_i]^+ \leq N + n$$

From (3.1.6) it follows that

$$L[M(\epsilon)] \geq L[M[\epsilon(N + n)]] > [N(N + n)] L[M(\hat{\epsilon})]$$

Introducing the change  $N + n = \tilde{N}$ , it is easy to obtain the validity of the theorem.

Equality in the right-hand side of (3.1.3) holds if and only if all  $N\hat{p}_i^*(i = 1, 2, \dots, n)$  are integers for at least one optimal design  $\hat{\epsilon}$ . In this case  $\hat{\epsilon}$ , and  $\hat{\epsilon}(N)$  coincide. In the other cases,  $\tilde{L}\{M[\hat{\epsilon}(N)]\} < \tilde{L}\{M(\hat{\epsilon})\}$ . As already noted earlier (cf. Section 1.10), the construction of optimal discrete designs  $\hat{\epsilon}(N)$  is a problem significantly more difficult than finding  $\hat{\epsilon}$ . To do this, in general, it is usually necessary to obtain a separate solution for each  $N$ . Therefore it is advantageous to find a procedure for correcting a continuous design on its approximated discrete design  $\hat{\epsilon}(N)$  which would according to the value of  $\tilde{L}\{M[\hat{\epsilon}(N)]\}$  deviate little from the design  $\hat{\epsilon}(N)$ .

One possible method is the following procedure: The design  $\hat{\epsilon}(N)$  is constructed by means of the allocation  $p_i \sim [(N - n)\hat{p}_i]^+$  ( $i = 1, 2, \dots, n$ ). For this design,  $\sum_{i=1}^n [(N - n)\hat{p}_i]^+ \leq N$ ; therefore the remaining unrealized observations  $N - \sum_{i=1}^n [(N - n)\hat{p}_i]^+$  can, for example, be added one-by-one up to the point where

$$(N - n)\hat{p}_j \geq [(N - n)\hat{p}_j]^+ - \frac{1}{2}.$$

The remainder  $N - \sum_{i=1}^n [(N - n)\hat{p}_i]^+$  can also be divided among the other points  $x \in X$ . We indicate the constructed design by the symbol  $\tilde{\epsilon}$ .

### Corollary 1

$$\tilde{L}\{M[\hat{\epsilon}(N)]\} - \tilde{L}\{M[\tilde{\epsilon}(N)]\} \leq (n/N)\tilde{L}\{M(\hat{\epsilon})\}. \quad (3.1.7)$$

Inequality (3.1.7) can be obtained if it is taken into account that

$$\begin{aligned} N\tilde{L}\{M[\tilde{\epsilon}(N)]\} &\geq \tilde{L}\left[\sum_{i=1}^n [(N - n)\hat{p}_i]^+ f(\hat{x}_i) f'(\hat{x}_i)\right] \\ &\geq \tilde{L}\left[\sum_{i=1}^n (N - n)\hat{p}_i f(\hat{x}_i) f'(\hat{x}_i)\right] = (N - n)\tilde{L}\{M(\hat{\epsilon})\}, \end{aligned}$$

or

$$\tilde{L}\{M[\hat{\epsilon}(N)]\} \geq \tilde{L}\{M[\tilde{\epsilon}(N)]\} \geq [(N - n)/N]\tilde{L}\{M(\hat{\epsilon})\}.$$

From this and from (3.1.3)

$$\tilde{L}\{M[\hat{\epsilon}(N)]\} - \tilde{L}\{M[\tilde{\epsilon}(N)]\} \leq (n/N)\tilde{L}\{M(\hat{\epsilon})\},$$

which is what was to be obtained.

If for the given problem there exist several continuous optimal designs then inequality (3.1.7) says that the most useful design to use is the design  $\epsilon$  the spectrum of which contains the smallest number of points. We exclude the rare case when  $\epsilon(N)$  and some continuous optimal design rounded off to discretize coincide [i.e. the numbers  $p_i N$  ( $i = 1, 2, \dots, n$ ) are integers]. Recall that for the majority of criteria of optimality the minimal number of points of the spectrum  $n$  for continuous optimal designs can be contained between the limits

$$m \leq n_0 < m(m+1)/2$$

**III** We will indicate some criteria for comparing designs which satisfy the requirements of (3.1.2) and (3.1.3)

1 Let the design  $\epsilon_1 > \epsilon_2$  if

$$M(\epsilon_1) > M(\epsilon_2)$$

Obviously the operation of taking the determinant does not satisfy the requirements (3.1.2) and (3.1.3). The given criterion of comparison can be readily transformed to an equivalent form

$$M(\epsilon_1)^{1/m} > M(\epsilon_2)^{1/m} \quad (3.1.8)$$

where  $m$  is the number of unknown parameters. The approach (3.1.8) satisfies the requirements (3.1.2) and (3.1.3). From (3.1.8) and (3.1.3) it is easy to obtain that the  $D$  optimal discrete designs and continuous ones are related by means of the relationship

$$[(N-n)/N]^m \cdot M(\epsilon) < M[\epsilon(N)] < |M(\epsilon)|$$

2 Let the design  $\epsilon_1 > \epsilon_2$  if

$$\max_x d(x, \epsilon_1) < \max_x d(x, \epsilon_2) \quad (3.1.9)$$

We transform criterion (3.1.9) to the equivalent form

$$\min_x d^{-1}(x, \epsilon_1) > \min_x d^{-1}(x, \epsilon_2)$$

We verify that the operator

$$L(A) = \min_x [f(x) A^{-1} f(x)]^{-1}$$

satisfies (3.1.2). Indeed, if  $A$  and  $B$  are positive-semidefinite matrices, then  $(A + B)^{-1} \leq A^{-1}$ , and it follows that

$$f'(x)(A + B)^{-1}f(x) \leq f'(x)A^{-1}f(x),$$

or, in agreement with the first condition of (3.1.2),

$$\min_x [f'(x)(A + B)^{-1}f(x)]^{-1} \geq \min_x [f'(x)A^{-1}f(x)]^{-1}.$$

The validity of the second condition of (3.1.2) follows from the fact that

$$f'(x)(kA)^{-1}f(x) = k^{-1}f'(x)A^{-1}f(x).$$

From this

$$\min_x [f'(x)(kA)^{-1}f(x)]^{-1} = \min_x k[f'(x)A^{-1}f(x)]^{-1}.$$

From (3.1.3) we have

$$[N/(N - n)] \max_x d(x, \hat{\epsilon}) > \max_x d[x, \epsilon(N)] \geq \max_x d(x, \hat{\epsilon}).$$

3. We return to the linear criteria for comparing designs. In particular, the  $A$ -criterion, the  $Q$ -criterion, and the criterion of equality of designs by extrapolation at a point are related to these criteria. Thus let  $\epsilon_1 > \epsilon_2$ , if

$$L[D(\epsilon_1)] < L[D(\epsilon_2)]$$

and the operator  $L$  satisfies the requirements (2.9.2)–(2.9.4). We set  $\tilde{L}[M(\epsilon)] = L^{-1}[M^{-1}(\epsilon)]$ . Since  $(A + B)^{-1} \leq A^{-1}$ , if  $A$  and  $B$  are positive-semidefinite matrices, then [cf. (2.9.2)]

$$L[(A + B)^{-1}] \leq L(A^{-1}),$$

and

$$\tilde{L}(A + B) = L^{-1}[(A + B)^{-1}] \geq L^{-1}(A^{-1}) = \tilde{L}(A).$$

That is, the first condition of (3.1.2) is satisfied. The second condition of (3.1.2) obviously follows from  $L(kA) = kL(A)$ .

Inequality (3.1.3) for the operator  $L$  can be rewritten in the form

$$[N/(N - n)] L[D(\hat{\epsilon})] > L\{D[\hat{\epsilon}(N)]\} \geq L[D(\hat{\epsilon})].$$

For example, for the  $A$ -criterion it is possible to write the chain of inequalities

$$[N/(N - n)] \operatorname{Tr} D(\hat{\epsilon}) > \operatorname{Tr} D[\hat{\epsilon}(N)] \geq \operatorname{Tr} D(\hat{\epsilon}).$$

IV. For the criteria for comparison of experiments considered in Part III we apply the method presented in Corollary 1 of Theorem 3.1.1 of rounding off continuous optimal designs to discrete designs. The obtained designs  $\hat{\epsilon}(N)$  will satisfy (3.1.7). For example, for  $D$  optimality the inequality (3.1.7) takes on the form

$$\{M[\hat{\epsilon}(N)]\} - \{M[\epsilon(N)]\} \leq \{1 - [(N-n)/N]^m\} M(\hat{\epsilon})^n$$

for the  $A$ -criterion

$$\text{Tr } D[\epsilon(N)] - \text{Tr } D[\hat{\epsilon}(N)] \leq [n/(N-n)] \text{Tr } D(\hat{\epsilon})$$

It is evident that the design  $\epsilon(N)$  is close to  $\hat{\epsilon}(N)$  when  $N$  is large and  $n$  is small. If  $N \gg n$ , then in the majority of the practical cases the laborious operation of seeking  $\hat{\epsilon}(N)$  can, without a real loss in accuracy of the obtained results, be transformed to a much simpler operation of rounding off a tabular design  $\epsilon$ . When  $N \approx n$ , in general we do not apply the indicated method of rounding off.

### 3.2. Properties and Methods of Constructing $D$ -Optimal Designs

In many investigations, particularly in experiments in real industrial settings, in view of the high cost of each observation, the number of measurements is kept to a minimum. In such cases, the search for discrete optimal designs becomes necessary.

Discrete designs do not observe many properties of continuous designs which are very useful for constructing the optimal designs. For example,  $D$ -optimal and minimax discrete designs are not equivalent for arbitrary  $N$  even when  $\lambda(x) = \text{const}$ . For arbitrary  $N$ , Theorem 2.9.2 is not satisfied. Moreover, the number of points in the spectrum of the optimal discrete design can exceed in general, the upper bound  $n = m(m+1)/2$  for the number of points in the spectrum of continuous optimal designs.

**EXAMPLE 1 [37]** We consider the linear regression on the interval

$$\eta(x|\theta) = \theta_1 + \theta_2 x, \quad \lambda(x) \equiv 1 \quad -1 \leq x \leq 1$$

Let  $N = 3$ . Relying on the obvious form of the dispersion matrix it is easy to verify that

$$\epsilon_1(3) = \left\{ \begin{array}{l} x_1 = -1, \quad x_2 = 0, \quad x_3 = 1 \\ p_1 = p_2 = p_3 = \frac{1}{3} \end{array} \right\}$$

minimizes  $\max_x d[x, \epsilon(3)]$ , and the design

$$\epsilon_2(3) = \begin{cases} x = -1, & x_2 = 1 \\ p_1 = \frac{1}{3}, & p_2 = \frac{2}{3} \end{cases}$$

minimizes the value of the determinant of the dispersion matrix. The indicated designs have the following characteristics:

$$|D[\epsilon_1(3)]| = \frac{3}{2},$$

$$\max_x d[x, \epsilon_1(3)] = d[1, \epsilon_1(3)] = d[-1, \epsilon_1(3)] = \frac{5}{2},$$

and

$$|D[\epsilon_2(3)]| = \frac{9}{8},$$

$$\max_x d[x, \epsilon_2(3)] = d[-1, \epsilon_2(3)] = 3.$$

**I.** In this part, our primary concern will be a specific investigation of some properties of discrete optimal designs [36], which permit us, later on, to develop sufficiently simple numerical methods for constructing optimal designs, i.e., simple in comparison with a direct search for the extrema of  $L\{M[\epsilon(N)]\}$  (for example, gradient or methods of random search).

We first consider  $D$ -optimal designs. For this it is helpful for us to operate with nonnormalized discrete designs which in distinction to normalized designs are indicated by  $\mathcal{E}(N)$ . Recall that  $M[\mathcal{E}(N)] = NM[\epsilon(N)]$ .

Let there be given a design  $\mathcal{E}(N)$  with spectrum  $x_1, x_2, \dots, x_n$  and let part of the measurements from the points  $x_{j_1}, x_{j_2}, \dots, x_{j_l}$ , belonging to this spectrum, be transformed to the arbitrary points  $x_k$  ( $k = 1, 2, \dots, l$ ). We indicate this new design by  $\tilde{\mathcal{E}}(N)$ .

**Lemma 3.2.1.** *Let  $M[\mathcal{E}(N)]$  be the information matrix of the design  $\mathcal{E}(N)$ ; then the determinant of the information matrix of the design  $\tilde{\mathcal{E}}(N)$  equals*

$$|M[\tilde{\mathcal{E}}(N)]| = |M[\mathcal{E}(N)]| \|I_l + F'M^{-1}[\mathcal{E}(N)]F|, \quad (3.2.1)$$

where

$$F = \|i\lambda^{1/2}(x_{j_1})f(x_{j_1}), \lambda^{1/2}(\tilde{x}_1)f(\tilde{x}_1), \dots, i\lambda^{1/2}(x_{j_l})f(x_{j_l}), \lambda^{1/2}(\tilde{x}_l)f(\tilde{x}_l)\|,$$

$$i = \sqrt{-1}.$$

*Proof.* By definition

$$M[\tilde{\mathcal{E}}(N)] = M[\mathcal{E}(N)] - \sum_{k=1}^l \lambda(x_k) f(x_k) f'(x_k) + \sum_{k=1}^l \lambda(\tilde{x}_k) f(\tilde{x}_k) f'(\tilde{x}_k)$$

This expression can be rewritten in the form  $M[\tilde{\mathcal{E}}(N)] = M[\mathcal{E}(N)] + FF'$ . Transforming with the last equality in Lemma 3.2.1, we obtain the necessary result

In what follows, we shall repeatedly run into the quantity  $|I + F'M^{-1}[\mathcal{E}(N)]F|$ . From the positive definiteness of the information matrices  $M[\mathcal{E}(N)]$  and  $M[\tilde{\mathcal{E}}(N)]$  and equality (3.2.1), it follows that this quantity is not less than zero and equality can hold only when  $|M[\mathcal{E}(N)]| = 0$

**Theorem 3.2.1.** At the points  $\tilde{x}_j$  ( $j = 1, 2, \dots, n$ ) in the spectrum of the discrete D-optimal design  $\mathcal{E}(N)$

$$\lambda(\tilde{x}_j) d(\tilde{x}_j) \geq \lambda(x) d(x) - \lambda(x) \lambda(\tilde{x}_j) [d(\tilde{x}_j) d(x) - d^2(\tilde{x}_j, x)], \quad (3.2.2)$$

where

$$\begin{aligned} d(x) &= d[x, \mathcal{E}(N)] = f(x) D[\mathcal{E}(N)] f(x), \\ d(x, \tilde{x}) &= d[x, \tilde{x}, \mathcal{E}(N)] = f(x) D[\mathcal{E}(N)] f(\tilde{x}), \end{aligned}$$

and  $x \in X$

*Proof* Let  $x$  be some point belonging to the region  $X$  of possible measurements. We assume that one of the measurements taken at the point  $\tilde{x}_j$ , belonging to the spectrum of the discrete D-optimal design, is transformed to this point. Utilizing the results of Lemma 3.2.1 we introduce the matrix

$$F = [\lambda^{1/2}(\tilde{x}_j) f(\tilde{x}_j), \lambda^{1/2}(x) f(x)]$$

In agreement with (3.2.1) the determinant of the information matrix of the new design  $\tilde{\mathcal{E}}(N)$  will be equal to

$$|M[\tilde{\mathcal{E}}(N)]| = |M[\mathcal{E}(N)]| |I + F M^{-1}[\mathcal{E}(N)] F|$$

Since

$$F M^{-1}[\mathcal{E}(N)] F = F D[\mathcal{E}(N)] F = \begin{vmatrix} -\lambda(\tilde{x}_j) d(\tilde{x}_j) & \lambda^{1/2}(\tilde{x}_j) \lambda^{1/2}(x) d(\tilde{x}_j, x) \\ \lambda^{1/2}(\tilde{x}_j) \lambda^{1/2}(x) d(\tilde{x}_j, x) & \lambda(x) d(x) \end{vmatrix}$$

and

$$\begin{aligned} |I + F' M^{-1} [\tilde{\mathcal{E}}(N)] F| &= 1 - \lambda(\hat{x}_j) d(\hat{x}_j) + \lambda(x) d(x) \\ &\quad - \lambda(x) \lambda(\hat{x}_j) [d(\hat{x}_j) d(x) - d^2(\hat{x}_j, x)] \end{aligned}$$

then

$$\begin{aligned} |M[\tilde{\mathcal{E}}(N)]| &= |M[\dot{\mathcal{E}}(N)]| \{1 - \lambda(\hat{x}_j) d(\hat{x}_j) + \lambda(x) d(x) \\ &\quad - \lambda(x) \lambda(\hat{x}_j) [d(\hat{x}_j) d(x) - d^2(\hat{x}_j, x)]\}. \end{aligned}$$

But by definition

$$|M[\tilde{\mathcal{E}}(N)]| \leq |M[\dot{\mathcal{E}}(N)]|,$$

so that from the two preceding relations it follows that

$$\lambda(\hat{x}_j) d(\hat{x}_j) \geq \lambda(x) d(x) - \lambda(\hat{x}_j) \lambda(x) [d(x) d(\hat{x}_j) - d^2(x, \hat{x}_j)].$$

The theorem is proved.

The last term in (3.2.2) is never less than zero. It is easy to show this by taking into account that  $d(\hat{x}_j, x)$  is the covariance of the two random variables  $\hat{\eta}(\hat{x}_j, \theta)$  and  $\hat{\eta}(x, \theta)$ , and that  $d(\hat{x}_j)$  and  $d(x)$  are the variances of these variables. It is well known (cf., for example, [12]) that

$$\rho^2(\hat{x}_j, x) = d^2(\hat{x}_j, x)/d(\hat{x}_j) d(x) \leq 1,$$

from which  $d(\hat{x}_j) d(x) - d^2(\hat{x}_j, x) \geq 0$  follows. In this way, at the points of the  $D$ -optimal discrete design the surface  $\lambda(x) d(x)$  need not attain its maximal value as in the continuous  $D$ -optimal designs. Theorem 3.2.1 is useful in verifying the  $D$ -optimality of designs. We now turn our attention to the fact that condition (3.2.2) is necessary but not sufficient.

Letting the number of possible observations approach infinity (i.e., passing to continuous designs) it may be shown that (3.2.2) with accuracy up to  $O(N^{-1})$  passes to the inequality

$$\lambda(\hat{x}_j) d(\hat{x}_j, \hat{\epsilon}) \geq \lambda(x) d(x, \hat{\epsilon}).$$

From this it also follows that

$$\lambda(\hat{x}_j) d(\hat{x}_j, \hat{\epsilon}) = \lambda(\hat{x}_k) d(\hat{x}_k, \hat{\epsilon}),$$

where  $\hat{x}_j$  and  $\hat{x}_k$  are distinct points of the spectrum of the design  $\hat{\epsilon}$ . The last two relations are in full correspondence with Theorem 2.2.1.

II. Relying on Theorem 3.2.1, it is possible to construct an iterative procedure for obtaining the spectrum of the  $D$ -optimal designs.

Let  $\mathcal{E}_0(N)$  be a design for which inequality (3.2.2) is not satisfied for points in its spectrum. We assume that the inequality is not satisfied for the point  $x_i$ . Transforming one of the measurements from the point  $x_i$  into the point  $x$  leads to an increase of the determinant of the information matrix

$$|M[\mathcal{E}_1(N)]| = |M[\mathcal{E}_0(N)]|(1 + \Delta_0(x_i, x))$$

if  $\Delta_0(x_i, x) > 0$ . Here

$$\Delta_0(x_i, x) = \lambda(x) d_s(x) - \lambda(x) \lambda(x_i) [d_s(x) d_s(x_i) - d_s^2(x, x_i)] - \lambda(x_i) d_s(x_i),$$

$$d_s(x) = d[x, \mathcal{E}_0(N)], \quad d_s(x, x_i) = d[x, x_i, \mathcal{E}_0(N)]$$

In order that the increase of the determinant of the information matrix for a given  $x_i$  be maximal at each stage, it is necessary to transform the measurement to the point  $\hat{x}$ , where  $\Delta_0(x_i, x)$  attains its maximal value. The increment of the determinant can be increased if we carry out a supplementary maximization in  $x_i$  ( $i = 1, 2, \dots, n_0$ ).

For the obtained designs  $\mathcal{E}_1(N)$ ,  $\max_i \max_x \Delta_1(x_i, x)$  ( $i = 1, 2, \dots, n_1$ ) is sought. If  $\max_i \max_x \Delta_1(x_i, x) > 0$  and the maximal value  $\Delta_1(x_i, x)$  is attained for  $i = j_1$  at the point  $\hat{x}$ , then one measurement is transformed from the point  $x_{j_1}$  into the point  $\hat{x}$ .

This procedure is continued until the following inequality holds

$$\max_i \max_x \Delta_i(x_i, x) \leq \delta, \quad (3.2.3)$$

where  $\delta$  is some small positive number specified beforehand. Since

$$|M[\mathcal{E}_0(N)]| \leq |M[\mathcal{E}_1(N)]| \leq \dots \leq |M[\mathcal{E}_k(N)]| \leq \dots \leq |M[\mathcal{E}(N)]|, \quad (3.2.4)$$

the convergence of the sequence  $\{|M[\mathcal{E}_k(N)]|\}$  follows from the existence of an upper bound.

The basic undesirable distinction of the procedure outlined from the analogous procedure for constructing continuous  $D$ -optimal designs is that the design to which the sequence  $\{\mathcal{E}_k(N)\}$  converges in some cases can be distinct from the optimal design

$$\lim_{k \rightarrow \infty} |M[\mathcal{E}_k(N)]| = |M[\mathcal{E}(N)]| \neq |M[\mathcal{E}_0(N)]|$$

Because of this, it is recommended that the iterative procedure be carried out several times beginning with various  $\mathcal{E}_0(N)$ . If the determinants of the information matrices all coincide, then the design  $\tilde{\mathcal{E}}(N)$  to which the corresponding sequences of designs  $\{\mathcal{E}_s(N)\}$  converges will, with large probability, coincide with the discrete optimal design. If the iterative procedures converge to different designs  $\tilde{\mathcal{E}}(N, j)$  ( $j = 1, 2, \dots, q$ ) for which the determinant of the information matrices are distinct, then it is recommended that trials be carried out, seeking the optimal design, until one can no longer extend the group of designs  $\tilde{\mathcal{E}}(N, j_1), \tilde{\mathcal{E}}(N, j_2), \dots, \tilde{\mathcal{E}}(N, j_l)$  ( $l < q$ ) with one for which the determinant of the information matrix exceeds all of the remaining.

In order to verify whether or not the design  $\tilde{\mathcal{E}}(N)$  is  $D$ -optimal it is also recommended that the relations (3.1.3) and (3.1.7) be used.

We note that in distinction from continuous  $D$ -optimal designs where all optimal designs for a given regression problem correspond to one and only one matrix  $M(\epsilon)$ , for each  $D$ -optimal discrete design there can exist in general a separate information matrix, but all determinants of these matrices are equal.

**EXAMPLE 2.** We consider the regression curve of the form

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3, \quad -1 \leq x \leq 1.$$

The efficiency function  $\lambda(x)$  will be assumed to be constant on the specified interval.

The continuous  $D$ -optimal design for the given regression problem is known and equals

$$\epsilon = \left\{ \begin{array}{l} x_1 = -1.0, \quad x_2 = -0.447, \quad x_3 = 0.447, \quad x_4 = 1.0 \\ p_i = 0.25 \quad (i = 1, 2, 3, 4) \end{array} \right\}.$$

According to the iterative procedure described in Section 3.2, discrete optimal designs were constructed for  $N = 4, 5, \dots, 15$  (cf. Fig. 17). It is clear that for  $N = 4, 8, 12$  the designs coincide according to the value of the determinant, normalized according to the number of measurements, with the value of the determinant of the continuous optimal designs. The remaining designs practically coincide with the rounded-off designs constructed according to the procedure outlined in Section 3.1 (cf. the remarks following Theorem 3.1.1).

**EXAMPLE 3.** In many cases, the number of points in the spectrum of the continuous  $D$ -optimal design significantly exceeds the number of

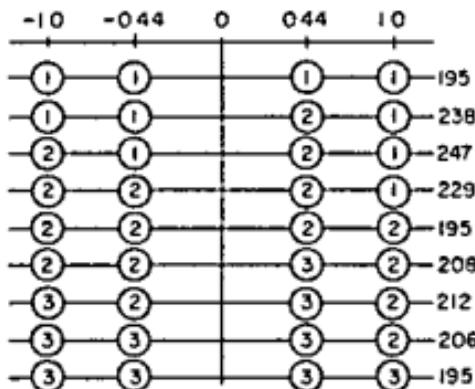


Fig. 17. Discrete optimal designs for cubic regression on an interval Successive rows are for increasing  $N$ . The right-hand column is  $N^{-1} |D(i)|$

sought parameters. In this connection, the problem arises of constructing designs which are sufficiently good in the sense of the value of the determinant of the dispersion matrix and at the same time consist of a small number of points. The development of the above iterative procedure permits the construction of such designs. Figure 18 shows other reduced designs with the number of points coinciding with the number of unknown parameters for the regression surfaces

$$\eta(x, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 + \theta_6 x_2^2$$

and

$$\eta(x, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 + \theta_6 x_2^2$$

$$+ \theta_7 x_1^2 x_2 + \theta_8 x_1 x_2^2 + \theta_9 x_1^3 + \theta_{10} x_2^3$$

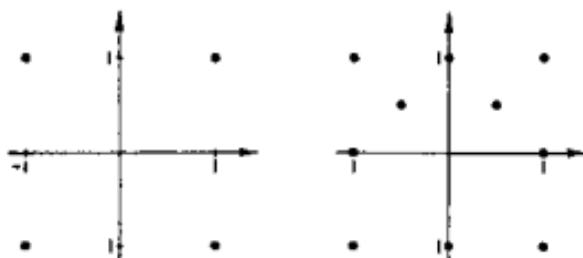


Fig. 18 Discrete optimal designs for polynomial regression of the second and third order for  $N = m$

It is evident that the specified designs are not unique. Indeed, "rotating" these designs by  $90^\circ$ , we again obtain designs with the same value of the determinant of the dispersion matrices.

### 3.3. Construction of Discrete Linear-Optimal Designs

I. We consider the properties of discrete linear optimal designs. For simplicity of exposition only nondegenerate designs will be considered in this section. The analysis of degenerate linear optimal designs depends on the device presented in Chapter 2 for the analysis of arbitrary continuous designs and essentially repeats the considerations presented in what follows.

We follow the notation introduced in the remarks preceding Lemma 3.2.1. We will prove the following two lemmas.

**Lemma 3.3.1.** *Let  $D[\mathcal{E}(N)]$  be the dispersion matrix of the design  $\mathcal{E}(N)$ ; then the dispersion matrix of the design  $\tilde{\mathcal{E}}(N)$  equals*

$$D[\tilde{\mathcal{E}}(N)] = [I_m - D[\mathcal{E}(N)]F\{I_l + F'D[\mathcal{E}(N)]F\}^{-1}F']D[\mathcal{E}(N)], \quad (3.3.1)$$

where  $I_q$  is the identity matrix of dimension  $q \times q$ ,  $m$  is the number of unknown parameters, and  $l/2$  is the number of translations  $x_{j_k} \rightarrow \tilde{x}_k$ .

*Proof.* Since

$$M[\tilde{\mathcal{E}}(N)] = M[\mathcal{E}(N)] + FF', \quad (3.3.2)$$

then, taking the inverse matrix on both sides of (3.3.2), we obtain

$$D[\tilde{\mathcal{E}}(N)] = \{I + D[\mathcal{E}(N)]FF'\}^{-1}D[\mathcal{E}(N)]. \quad (3.3.3)$$

Setting  $A = D[\mathcal{E}(N)]F$  and  $B = F'$  and using Lemma 2.6.1 we obtain

$$\{I_m + D[\mathcal{E}(N)]FF'\}^{-1} = I_m - D[\mathcal{E}(N)]F\{I_l + F'D[\mathcal{E}(N)]F\}^{-1}F'. \quad (3.3.4)$$

Combining (3.3.3) and (3.3.4) we succeed in verifying the lemma.

**Lemma 3.3.2.** *Let the operator  $L$  have the properties (2.9.2); then*

$$L\{D[\tilde{\mathcal{E}}(N)]\} = L\{D[\mathcal{E}(N)]\} \\ - L(D[\mathcal{E}(N)]F\{I_l + F'D[\mathcal{E}(N)]F\}^{-1}F'D[\mathcal{E}(N)])]. \quad (3.3.5)$$

*Proof* In order to show the validity of (3.3.5) it is sufficient to act on both sides of equation (3.3.1) with the operator  $L$  and make use of property (2.9.2).

We will show that discrete linear optimal designs satisfy properties analogous to those given in the exposition in the preceding section for  $D$  optimal designs.

**Theorem 3.3.1.** *For the points  $\hat{x}_j$  ( $j = 1, 2, \dots, n$ ) of the spectrum of the discrete linear optimal design  $\mathcal{E}(N)$*

$$\begin{aligned} [1 + \lambda(x) d(x)] \varphi(\hat{x}_j) &\geq [1 - \lambda(\hat{x}_j) d(\hat{x}_j)] \varphi(x) \\ &+ \lambda^{1/2}(\hat{x}_j) \lambda^{1/2}(x) d(\hat{x}_j, x) [\varphi(x, \hat{x}_j) + \varphi(\hat{x}_j, x)] \end{aligned} \quad (3.3.6)$$

where  $x \in X$ ,

$$\begin{aligned} \varphi(x) &= \varphi[x, \mathcal{E}(N)] = \lambda(x) L\{D[\mathcal{E}(N)] f(x) f'(x) D[\mathcal{E}(N)]\}, \\ \varphi(x, \hat{x}_j) &= \varphi[x, \hat{x}_j, \mathcal{E}(N)] = \lambda^{1/2}(x) \lambda^{1/2}(x) L\{D[\mathcal{E}(N)] f(x) f'(x) D[\mathcal{E}(N)]\} \end{aligned}$$

The remaining notation is the same as in Theorem 3.2.1.

*Proof* Let  $x$  be some point belonging to the region  $X$ , the set of possible measurements, and let one measurement be transformed from the point  $\hat{x}_j$  into the point  $x$ , then

$$F = \|\lambda^{1/2}(\hat{x}_j) f(\hat{x}_j) \lambda^{1/2}(x) f(x)\| \quad (3.3.7)$$

We denote by  $\mathcal{E}'(N)$  again the obtained design. Setting (3.3.7) into (3.3.5) it is easy to show that

$$\begin{aligned} D(\hat{x}_j, x) &= L\{D[\mathcal{E}'(N)]\} - L\{D[\mathcal{E}(N)]\} \\ &= |I_2 + F D[\mathcal{E}'(N)] F|^{\frac{1}{2}} \{[1 - \lambda(\hat{x}_j) d(\hat{x}_j)] \varphi(x) \\ &+ \lambda(\hat{x}_j) \lambda(x) d(\hat{x}_j, x) [\varphi(x, \hat{x}_j) + \varphi(\hat{x}_j, x)] \\ &- [1 - \lambda(x) d(x)] \varphi(x)\} \end{aligned} \quad (3.3.8)$$

Since the design  $\mathcal{E}'(N)$  is optimal and the determinant

$$|I_2 + F D[\mathcal{E}'(N)] F|$$

is greater than zero (cf. the remarks to Lemma 3.2.1), then

$$L\{D[\mathcal{E}'(N)]\} - L\{D[\mathcal{E}(N)]\} \leq 0 \quad (3.3.9)$$

Combining (3.3.8) and (3.3.9) we obtain the validity of (3.3.6), and the theorem is proved.

For  $A$ -optimal discrete designs inequality (3.2.6) has the form

$$\begin{aligned} [1 + \lambda(x) d(x)] \varphi(\hat{x}_j) &\geq [1 - \lambda(\hat{x}_j) d(\hat{x}_j)] \varphi(x) \\ &+ 2\lambda^{1/2}(x) \lambda^{1/2}(\hat{x}_j) d(\hat{x}_j, x) \varphi(\hat{x}_j, x), \end{aligned} \quad (3.3.10)$$

where

$$\begin{aligned} \varphi(x) &= \lambda(x) f'(x) D^2[\mathcal{E}(N)] f(x), \\ \varphi(x, \tilde{x}) &= \varphi(\tilde{x}, x) = \lambda^{1/2}(x) \lambda^{1/2}(\tilde{x}) f'(x) D^2[\mathcal{E}(N)] f(\tilde{x}). \end{aligned}$$

If we go to normalized quantities, then (3.3.10) can be rewritten with accuracy up to  $O(N^{-1})$  in the form

$$[1 + O(N^{-1})] \varphi(\hat{x}_j, \epsilon(N)) \geq [1 - O(N^{-1})] \varphi(x, \epsilon(N)) + O(N^{-1}). \quad (3.3.11)$$

As  $N \rightarrow \infty$  inequality (3.3.11) transforms into a familiar inequality (cf. Section 2.11), holding for continuous  $A$ -optimal designs:

$$\lambda(\hat{x}_j) f'(\hat{x}_j) D^2(\epsilon) f(\hat{x}_j) \geq \lambda(x) f'(x) D^2(\epsilon) f(x). \quad (3.3.12)$$

**II.** For numerical construction of optimal linear designs as in the construction of  $D$ -optimal discrete designs, it is useful to make use of iterative methods.

The iterative procedure consists of the following four operations.

1. Let there be some nondegenerate design  $\mathcal{E}_s(N)$  with dispersion matrix  $D[\mathcal{E}_s(N)]$ . The points  $x_{j_s}$  and  $\hat{x}$  are sought corresponding to

$$\max_i \max_x \Delta_s(x_i, x), \quad i = 1, 2, \dots, n_s,$$

where  $\Delta_s(x_i, x)$  is defined by means of formula (3.3.8).

2. One measurement is transformed from the point  $x_{j_s}$  into the point  $\hat{x}$ . The obtained design will be indicated by  $\mathcal{E}_{s+1}(N)$ .

3. The dispersion matrix is computed:

$$D[\mathcal{E}_{s+1}(N)] = [I_m - D[\mathcal{E}_s(N)] F \{I_2 + F' D[\mathcal{E}_s(N)] F\}^{-1} F'] D[\mathcal{E}_s(N)].$$

4. Operations 1–3 are repeated with the index  $s$  changed to  $s + 1$ , and so on.

The computation is terminated if

$$\max_i \max_x D_s(x_i, x) \leq \delta, \quad (3.3.13)$$

where  $\delta$  is a previously given "accuracy." Since

$$L[D(\mathcal{E}_0(N))] \geq L[D(\mathcal{E}_1(N))] \geq \dots \geq L[D(\mathcal{E}_s(N))],$$

then the given sequence converges in view of the fact that  $L[D(\mathcal{E}_s(N))] \geq L[D(\mathcal{E}(N))]$ , where  $\mathcal{E}(N)$  is a linear optimal design. However, in some cases the limit of this sequence can be strictly greater than  $L[D(\mathcal{E}(N))]$ . Therefore, just as in the iterative procedure for seeking discrete  $D$ -optimal designs it is recommended that the iterative procedure be repeated with several distinct starting designs.

III. Despite the cumbersome expressions for  $D_s(x_i, x)$  the iterative procedures just described for seeking  $D$ - and  $L$ -optimal designs are significantly simpler, in the majority of cases, than directly seeking  $\max_{x(N)} |M[\mathcal{E}(N)]|$  or  $\min_{x(N)} L[D(\mathcal{E}(N))]$  when an electronic computing machine is available. Indeed, any direct method of seeking the conditional extremum (i.e., this can be one of the modifications of gradient methods, methods of descent, etc.) is extremely sensitive to the growth of the dimension of the space in which the extremum is sought. For example, for one of the simple regression problems, linear regression in a factor space of dimension  $k$ , the search for the conditional extremum must be carried out in a space of at least dimension  $(k+1)k+k$  variables, and  $k$  of these are in discrete form. In this process of seeking the extremum, which usually has an iterative character, it is necessary at each step to compute either  $|M[\mathcal{E}(N)]|$  or to invert the matrix  $M[\mathcal{E}(N)]$  of dimension  $m \times m$ , where  $m$  is the number of unknown parameters.

In the iterative procedure presented in this part, the search for the extremum is carried out in a space of  $k+1$  variables (one of the variables  $x_i$  is discrete). The search of  $\max_i \max_x D_s(x_i, x)$  can be carried out not to the end, but stopped as soon as  $D_s(x_i, x)$  exceeds some given level. This somewhat increases the number of iterations but significantly reduces the volume of the average computation. In the considered algorithm, we succeeded in avoiding the computation of determinants and the inversion of matrices of large dimension. The nature of the iterative procedures is such that these operations are transformed to operations of calculating quadratic forms  $f D f$ .

# **4**

## **Sequential Methods of Designing Experiments for Refining and Determining Estimates of the Parameters**

### **4.1. Some Generalities of Contemporary Experimental Investigations**

In contemporary experimental practice it is necessary to deal with experiments complicated from the point of view of technically carrying them out and also from the point of view of their theoretical interpretation. Many of these experiments are of such a duration that during the time they are conducted the experimental conditions change in an essential way. What, generally speaking, is understood by these conditions?

As a rule no fundamental investigation can be considered as isolated from the development of a given branch of science or technology. Often several groups of people work on the same problem, conducting theoretical and experimental investigations, if not overlapping, then complementing one another. Therefore, in the course of investigation, an experimenter, besides having information obtained in a particular laboratory, also has supplementary information either due to direct contact with related laboratories, or from literature sources. This information can take on a varied character. It can be theoretical

results, refining the basic form of the changing analytic form of the response surface  $\eta(x, \theta)$ , new results concerning experimental methods, the appearance of new technology, more current apparatus and setups, and finally immediate experimental data about the investigated variables.

As one example of such a complicated and long investigation, we may take an experiment in the investigation of the interaction of elementary particles. In particular, in the Joint Institute of Nuclear Research, the first experiments on the interaction of nucleons with energy 660 MeV were carried out before 1950 (cf [38]) and continued up to 1966 (cf [39, 40]). Surveying the cited works, one may without difficulty track down the influence of all the factors mentioned above in the character of the experimental investigation in a given narrow branch of the physics of elementary particles.

It is evident that under conditions analogous to those presented above, it makes no sense to construct a design which would specify the distribution of all resources allotted for the given investigations (statistical design). Clearly this design would usually be nonoptimal, since it is impossible beforehand to predict all possible situations (in this respect simplification in quantitative characteristics) awaiting the experimenter on his long and difficult path to the goal.

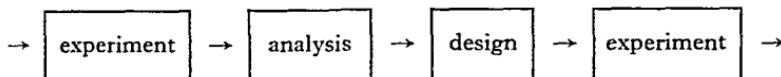
The majority of contemporary investigations are begun with a careful selection of information, available in the world's scientific literature or other sources. In other words, the experiment begins not in a vacuum, where information about the sought parameters (response surface) is absent, but in the presence of some prior information. Statistical methods are comparatively easily realized only in the absence of initial information [ $M(0) = 0$ ]. In the other case [ $M(0) \neq 0$ ] the amount of mathematical computations is most often so large that such designs become significantly more cumbersome than the experiment itself.

In this way, for these cases, the design of experiments, using the statistical methods of the previous chapters, must be ruled out. In this case it is more expedient to turn to so-called sequential methods of designing experiments in the determination and refinement of unknown parameters. These methods are also very useful for the sequential procedure considered in the introduction (cf. Diagram 1, page 8).

The idea of a sequential design consists of the following. The resources (for example, time) are divided into small "portions." The

experiment is divided into several steps and at each step a design is carried out, using one "portion" of the resources. An analysis of the experiment is conducted after each step.

The scheme of the entire process of the experiment will have the following form:



The analysis block is understood to be not only the usual regression analysis of experimental data, but also analysis of information occurring externally. The experiment ceases as soon as a given characteristic of precision of the estimates of the group of parameters  $\theta$  attains a prescribed value (for example, the determinant of the covariance matrix of the estimates of the sought parameters  $|D(\hat{\theta})|$ , the sum of its diagonal elements, etc.).

#### 4.2. Sequential D-Optimal Design (Linear Parametrization and Time Constant Efficiency of the Experiment)

I. In this chapter, it will be useful for us to use nonnormalized designs. Let the results of the experiments be compared according to the value of the determinant of the matrix

$$D_u = \begin{vmatrix} D_{11} & D_{12} & \cdots & D_{1l} \\ D_{21} & D_{22} & \cdots & D_{2l} \\ \vdots & \vdots & & \vdots \\ D_{u1} & D_{u2} & \cdots & D_{ul} \end{vmatrix},$$

where  $D_{\alpha\beta}$  ( $\alpha, \beta \leq l$ ) are the elements of the covariance matrix, corresponding to the parameters of interest to the experimenter. The design  $\mathcal{E}_1(T)$  will be preferred to the design  $\mathcal{E}_2(T)$  if for the same resources

$$|D_u[\mathcal{E}_1(T)]| < |D_u[\mathcal{E}_2(T)]|, \quad (4.2.1)$$

where  $T$  denotes the resources allocated to the given experiment. We will assume that the cost of each observation does not depend on when and where (in the space  $X$ ) the measurement is taken. Then  $T = cN$ , where  $c$  is the cost of one observation and  $N$  is the number

of observations allocated for the current experiment Inequality (4.2.1) can then be rewritten in the form

$$|D_{ii}[\mathcal{E}_1(N)]| < |D_{ii}[\mathcal{E}_2(N)]| \quad (4.2.2)$$

We will study the properties of experiments conducted in the following manner

1 At each given moment of time (which we will characterize by means of the number of observations  $N$ ) a measurement is taken at that point where the possible decrease of the determinant  $|D_{ii}|$  is maximal

2 After each observation is taken, an analysis of the obtained data is conducted Operation 1 is then repeated, and so forth

The aim of the current section is to derive formulas that will simplify the computational procedures for the sequential method of design presented and will qualify the usefulness of such a design (in other words, it will clarify whether or not the strategy for conducting the experiment is optimal, and if optimal, then under what conditions)

II From the computational viewpoint, sequential designs constructed using the criterion (4.2.2) consist of finding at each stage

$$\min_x |D_{ii}(N+1, x)| = \min_x |D_{ii}[\mathcal{E}(N+1, x)]| \quad (4.2.3)$$

Since the manifest expression giving the characteristics of the response surface, depends only on the information matrix  $M(N+1, x)$  (cf Chapter 1) then for finding the minimum (4.2.3) it is necessary to carry out the following operations

- 1 Construct the matrix  $M(N+1, x)$
- 2 Compute the inverse of this matrix  $D(N+1, x) = M^{-1}(N+1, x)$
- 3 Compute the determinant  $|D_{ii}(N+1, x)|$

The most cumbersome and least accurate steps in carrying out the computations on a calculating machine are the inversion of matrices and the computation of determinants If it is taken into account that for the computation of the minimum (for example, according to one of the modified gradient methods) many repetitions of these operations are required, then the necessity of a formula which would permit us to replace the inversion of the matrix and the computation of the determinant by more simple operations becomes evident

Initially we consider the case when the function  $\eta(x, \theta)$  depends linearly on the true parameters  $\theta$  [41, 42]:

$$\eta(x, \theta) = \sum_{\alpha=1}^m \theta_\alpha f_\alpha(x),$$

and the region of experimentation  $X$ , the efficiency  $\lambda(x)$ , and the functions  $f_\alpha(x)$  ( $\alpha = 1, 2, \dots, m$ ) satisfy the same conditions as in Chapter 2, Section 2.2. We assume that after  $N$  observations the information matrix has the value  $M(N)$ . At the point  $x$  let  $\Delta N$  supplementary measurements be taken. We find the relationship between the elements of the matrix  $D(N) = M^{-1}(N)$  and  $D(N + \Delta N) = M^{-1}(N + \Delta N)$  and the determinants  $|D_{ll}(N)|$  and  $|D_{ll}(N + \Delta N, x)|$ .

**Lemma 4.2.1.** [41].

$$D(N + \Delta N, x) = \left( I - \frac{\lambda(x) \Delta N D(N) f(x) f'(x)}{1 + \lambda(x) \Delta N d(x, N)} \right) D(N), \quad (4.2.4)$$

where  $d(x, N) = f'(x) D(N) f(x)$  is the variance of the response surface at the point  $x$  after  $N$  observations.

*Proof.* By definition (cf. Chapter 1),

$$M(N + \Delta N, x) = M(N) + \lambda(x) \Delta N f(x) f'(x),$$

or

$$\begin{aligned} D(N + \Delta N, x) &= [M(N) + \lambda(x) \Delta N f(x) f'(x)]^{-1} \\ &= [I_m + \lambda(x) \Delta N D(N) f(x) f'(x)]^{-1} D(N). \end{aligned} \quad (4.2.5)$$

Setting  $A = \lambda(x) \Delta N D(N) f(x)$  and  $B = f'(x)$  we find, from (4.2.5) and Lemma 2.6.1, that

$$D(N + \Delta N, x) = \left[ I_m - \frac{\lambda(x) \Delta N D(N) f(x) f'(x)}{1 + \lambda(x) \Delta N d(x, N)} \right] D(N),$$

which was required to be shown.

**Lemma 4.2.2.** [42].

$$|D_{ll}(N + \Delta N, x)| = |D_{ll}(N)| \left[ 1 - \frac{\lambda(x) \Delta N q'(x) D_{ll}^{-1}(N) q(x)}{1 + \lambda(x) \Delta N d(x, N)} \right], \quad (4.2.6)$$

where

$$q_\alpha(x) = \sum_{\beta=1}^m D_{\alpha\beta}(N) f_\beta(x), \quad \alpha = 1, 2, \dots, l$$

*Proof* From Lemma 4.2.1 it follows that

$$|D_{ll}(N + \Delta N, x)| = \left| \left\| D(N) - \frac{\lambda(x) \Delta N D(N) f(x) f(x)^T D(N)}{1 + \lambda(x) \Delta N d(x, N)} \right\|_1^l \right| \quad (4.2.7)$$

We set

$$D = 1, \quad A = D_{ll}(N), \quad B = \left\| \frac{\lambda(x) \Delta N D(N) f(x)}{1 + \lambda(x) \Delta N d(x, N)} \right\|_1^l,$$

and

$$C = \|f(x) D(N)\|_1^l$$

Then from (4.2.7) and Theorem 1.13 we obtain

$$|D_{ll}(N + \Delta N, x)|$$

$$= |D_{ll}(N)| \left[ 1 - \frac{\lambda(x) \Delta N}{1 + \lambda(x) \Delta N d(x, N)} \|f(x) D(N)\|_1^l D_{ll}^{-1}(N) \|D(N) f(x)\|_1^l \right].$$

or introducing the notation  $q_\alpha(x) = \sum_{\beta=1}^m D_{\alpha\beta}(N) f_\beta(x)$  we have

$$|D_{ll}(N + \Delta N, x)| = |D_{ll}(N)| \left[ 1 - \frac{\lambda(x) \Delta N q(x) D_{ll}^{-1}(N) q(x)}{1 + \lambda(x) \Delta N d(x, N)} \right]$$

which is what was required to be shown

In using the results of Lemmas 4.2.1 and 4.2.2 the operation of computing the determinant is carried out only once in obtaining  $\min_x |D_{ll}(N + \Delta N, x)|$  ( $|D_{ll}(0)|$  is computed, where  $D(0)$  is the value of the covariance matrix at the initial design of the experiment). The inversion of a matrix is carried out once at each step of the design [ $D_{ll}^{-1}(N)$  is computed]. In this manner Lemmas 4.2.1 and 4.2.2 permit us to describe the design process in terms of the following system

$$\frac{\Delta N \lambda(x_N) q(x_N) D_{ll}^{-1}(N) q(x_N)}{1 + \Delta N \lambda(x_N) d(x_N, N)} = \max_x \frac{\Delta N \lambda(x) q(x) D_{ll}^{-1}(N) q(x)}{1 + \Delta N \lambda(x) d(x, N)} \quad (4.2.8)$$

$$D(N+1) = \left[ I_m - \frac{\Delta N \lambda(x_N) D(N) f(x_N) f(x_N)^T D(N)}{1 + \Delta N \lambda(x_N) d(x_N, N)} \right] D(N)$$

In system (4.2.8) it is assumed for simplicity that  $\Delta N = 1$ , although it is possible to consider the sequential procedure also for  $\Delta N > 1$ .

A particularly simple and obvious form is taken on by (4.2.6) for  $l = m$  and  $l = 1$ :

$$|D(N + \Delta N, x)| = [1 + \lambda(x) \Delta N d(x, N)]^{-1} |D(N)|, \quad l = m, \quad (4.2.9)$$

and

$$D_{\alpha\alpha}(N + \Delta N, x) = D_{\alpha\alpha}(N) \left[ 1 - \frac{\lambda(x) \Delta N [\sum_{\beta=1}^m D_{\alpha\beta}(N) f_{\beta}(x)]^2 D_{\alpha\alpha}^{-1}(N)}{1 + \lambda(x) \Delta N d(x, N)} \right], \\ l = 1. \quad (4.2.10)$$

In these cases, the operations of inversion of the matrix  $[D(N)]$  is computed] are carried out once at the first step of the sequential design.

We stop with the case of refining estimates of all parameters  $l = m$ . The first equation of system (4.2.8) in this case becomes equivalent to the equation

$$\lambda(x_N) d(x_N, N) = \max_x \lambda(x) d(x, N). \quad (4.2.11)$$

Equation (4.2.11) has the same form for any  $\Delta N$ . It follows that the position of the point  $x_N$  at which it is necessary to take  $\Delta N$  observations does not depend on the value  $\Delta N$ . The position of this point also does not change if the efficiency function  $\lambda(x)$  is multiplied by any positive constant. This fact is of utmost importance in the practical application of (4.2.8), since in many cases the efficiency function  $\lambda(x)$  is known only with accuracy up to a constant multiplier (cf., e.g., [43]). We note the following interesting property of the sequential design for  $l = m$ . Equation (4.2.11) specifies that at each moment of time the observation must be taken at a point where the variance of the estimate of the response surface is maximal [for simplicity we consider the case  $\lambda(x) \equiv \text{const}$ ]. In other words, the experimenter can obtain a maximum amount of information about the surface being investigated from the observations at those points where the least is known about it.

III. Sequential designs depending on the system (4.2.8) give the maximal increment of information (in the sense of the value of the determinant  $|D_{ll}|$ ) at each distinct step. But this in general does not indicate that the overall strategy of the experiment is optimal in the large.

We consider at first the case when the efficiency  $\lambda(x)$  does not change in time (e.g., the experimenter works with one and the same apparatus, with quantities of one and the same degree of purity, etc.). In this case (cf. Chapter 2) there exists a statistically  $D$  optimal continuous design  $\mathcal{E}(N)$ . We indicate the covariance matrix of the estimate of the parameters  $\theta_1, \theta_2, \dots, \theta_l$  corresponding to the normalized  $D$  optimal design by  $D_{1l}(\hat{\epsilon})$ .

It is obvious that for any design  $\mathcal{E}(N)$  the following inequality must be satisfied

$$N^{-1} |D_{1l}(\epsilon)| \leq |D_{1l}(\mathcal{E}(N))| \quad (4.2.12)$$

The smaller the difference between  $N^{-1} |D_{1l}(\hat{\epsilon})|$  and  $|D_{1l}(\mathcal{E}(N))|$  the better we will consider the sequential design. By  $D(N)$  is understood the covariance matrix of the corresponding sequential design  $\mathcal{E}(N)$ .

The following theorem is valid.

**Theorem 4.2.1** *The sequential design carried out according to the strategy presented above is asymptotically (as  $N \rightarrow \infty$ )  $D$  optimal.*

$$\lim_{N \rightarrow \infty} N^{-1} |D_{1l}(N)| = |D_{1l}(\hat{\epsilon})|$$

*Proof.* In order to minimize the algebraic computations we consider the case  $l = m$ . For truncated  $D$  optimal designs the proof is constructed by completely analogous methods.

We pass to normalized designs. Then the system (4.2.8) takes on the form

$$\lambda(x_N) d[x_N - \epsilon(N)] = \max_x \lambda(x) d[x - \epsilon(N)] \quad (4.2.13)$$

$$(N+1)^{-1} D[\epsilon(N+1)] = \left[ I_m - \frac{N^{-1} \lambda(x_N) D[\epsilon(N)] f(x_N) f'(x_N)}{1 + N^{-1} \lambda(x_N) d[x_N - \epsilon(N)]} \right] N^{-1} D[\epsilon(N)]$$

where  $\epsilon(N)$  is the normalized design corresponding to the design  $\mathcal{E}(N)$ . We set  $\alpha_N = (N+1)^{-1}$ . Then (4.2.13) can be rewritten in the form

$$\lambda(x_N) d[x_N - \epsilon(N)] = \max_x \lambda(x) d[x - \epsilon(N)] \quad (4.2.14)$$

$D[\epsilon(N+1)]$

$$= (1 - \alpha_N)^{-1} \left[ I_m - \frac{\alpha_N \lambda(x_N) D[\epsilon(N)] f(x_N) f'(x_N)}{1 - \alpha_N + \alpha_N \lambda(x_N) d[x_N - \epsilon(N)]} \right] D[\epsilon(N)]$$

or, as  $N \rightarrow \infty$ , keeping in (4.2.14) terms of order no larger than  $\alpha_N$ , we obtain

$$\begin{aligned} \lambda(x_N) d[x_N, \epsilon(N)] &= \max_x \lambda(x) d[x, \epsilon(N)], \\ D[\epsilon(N+1)] &= D[\epsilon(N)] - \alpha_N D[\epsilon(N)] f(x_N) f'(x_N) D[\epsilon(N)]. \end{aligned} \quad (4.2.15)$$

Comparing system (4.2.15) with the iterative procedure of constructing  $D$ -optimal designs considered in Section 2.5, it is not difficult to obtain the asymptotic equivalence of normalized designs corresponding to the designs  $\epsilon(N)$ , with the designs  $\epsilon_N$  which are constructed at each  $N$ th iteration of the method presented in Section 2.5, for  $\alpha_N = (N+1)^{-1}$ . But by Theorem 2.5.3 (see also the explanation of this theorem, presented in the conclusion of Section 2.6)

$$\lim_{N \rightarrow \infty} |D(\epsilon_N)| = D(\hat{\epsilon}).$$

It follows that

$$\lim_{N \rightarrow \infty} |D[\epsilon(N)]| = \lim_{N \rightarrow \infty} N^m |D(N)| = |D(\hat{\epsilon})|,$$

which is what was required to be shown.

In practical investigations, for asymptotically optimal strategies it is very useful to know how far the sequential designs  $\epsilon(N)$  differ from the optimal design  $\hat{\epsilon}$ . In connection with this we investigate the question of an upper bound for

$$|D_u(N)| = N^{-l} |D_u[\epsilon(N)]|.$$

We assume that at the beginning of the experiment, to be conducted according to the strategy presented in the preceding Part II, the covariance matrix of the estimates of the parameters has the value  $D(N_0)$ . If the design  $\hat{\epsilon}$  is a continuous  $D$ -optimal design for a given regression problem, then there is a number  $\tilde{N}$ , such that

$$\tilde{N}^{-l} |D_u(\hat{\epsilon})| = |D_u[\mathcal{E}(N_0)]|, \quad (4.2.16)$$

where  $\tilde{N} + c = N_0$ ,  $c > 0$ , and  $N_0$  is the number of observations, allocated to the preliminary experiment, for which the estimates of the unknown parameters have covariance matrix  $D[\mathcal{E}(N_0)]$ . The quantity  $c$  can be treated as a loss following the nonoptimal preliminary experiment.

We consider the case when  $l = m$ . From (4.2.9) it follows that after one observation the logarithm of the determinant of the dispersion matrix decreases by the quantity

$$\begin{aligned}\Delta_N &= \log |D(N)| - \log \frac{|D(N)|}{1 + \lambda(x_N) d(x_N, N)} \\ &= \log \{1 + N^{-1} \lambda(x_N) d[x_N, \epsilon(N)]\}\end{aligned}$$

By Theorem 2.2.2  $\max_x \lambda(x) d[x, \epsilon(N)] > m$ , if the design  $\epsilon(N)$  is not  $D$ -optimal. Therefore

$$\Delta_N \geq \log \left(1 + \frac{m}{N}\right) \quad (4.2.17)$$

On the other hand, for a continuous  $D$ -optimal design, satisfying condition (4.2.16), the logarithm of the determinant of the covariance matrix, as the allocation increases by one, decreases by the quantity

$$\begin{aligned}\Delta_N &= \log(N - c)^{-m} |D(\hat{\epsilon})| - \log(N - c + 1)^{-m} |D(\hat{\epsilon})| \\ &= \log \{1 + 1/(N - c)\}^m \quad (4.2.18)\end{aligned}$$

Relying on formulas (4.2.17) and (4.2.18) it is not difficult to prove the following assertion

**Theorem 4.2.2.** *If, for the sequential design conducted according to the strategy presented above with  $N = N_0$ , the following inequality holds*

$$1 + (m/N_0) \geq [1 + (N_0 - c)^{-1}]^m \quad (4.2.19)$$

then

$$|D(N)| \leq (N - c)^{-m} |D(\hat{\epsilon})| \quad (4.2.20)$$

*Proof.* Both sides of inequality (4.2.20) are functions which are equal for  $N = N_0$  and monotone increasing as  $N$  increases. The rate of their growth is determined by formulas (4.2.17) and (4.2.18), respectively. Therefore, for the proof of the theorem it is sufficient to show that

$$\log[1 + (m/N)] \geq \log[1 + (N - c)^{-1}]^m \quad (4.2.21)$$

for any  $N \geq N_0$ .

Considering  $N$  as a continuous variable and differentiating both sides of inequality (4.2.21) with respect  $N$ , we obtain

$$\begin{aligned} (\partial/\partial N) \log[1 + (m/N)] &= -m/(N + m) N \\ &\geq (\partial/\partial N) \log[1 + (N - c)^{-1}]^m \\ &= -m/[(N - c)(N - c + 1)], \end{aligned} \quad (4.2.22)$$

in which case inequality holds only for  $c = 0$  and  $m = 1$ . From (4.2.19) and (4.2.22) the validity of inequality (4.2.21) obviously follows. The theorem is proved.

Generalization of Theorem 4.2.2 to the case  $l < m$ , although somewhat cumbersome from the computational point of view, presents no principal difficulties and can be carried out by the reader independently.

If  $N_0 \gg l$ , then instead of Theorem 4.2.2 it is useful to use its asymptotic analog [44].

**Theorem 4.2.2a.** *If  $N_0 \gg l$ , then the following inequality is valid:*

$$|D_{il}(N)| \leq (N - c)^{-l} |D_{il}(\hat{\epsilon})|. \quad (4.2.23)$$

*Proof.* As in Theorem 4.2.1 the proof is conducted only for the case  $l = m$ . For sufficiently large  $N_0$  the quantities  $|D(N)|$  and  $N^{-m} |D(\hat{\epsilon})|$  can be considered as functions of some continuous parameter  $N$ . Then

$$(\partial/\partial N) \log(N - c)^{-m} |D(\hat{\epsilon})| = -m/(N - c),$$

and [cf. (4.2.9)]

$$(\partial/\partial N) \log |D(N)| = -\lambda(x_N) d(x_N, N) = -N^{-1} \lambda(x_N) d[x, \epsilon(N)].$$

In view of the equivalence theorem (cf. Section 2.2)

$$\max_x \lambda(x) d[x, \epsilon(N)] = \lambda(x_N) d[x_N, \epsilon(N)] > m.$$

It follows that

$$(\partial/\partial N) \log |D(N)| < -m/N,$$

or, for any  $N \geq N_0$ ,

$$(\partial/\partial N) \log |D(N)| < (\partial/\partial N) \log(N - c)^m |D(\hat{\epsilon})|. \quad (4.2.24)$$

Since [cf condition (4.2.16)]

$$\log |D(N_0)| = \log N^{-m} |D(\epsilon)|$$

then from (4.2.24) and the boundedness of the corresponding second derivative it follows that

$$\log |D(N)| < \log(N - \epsilon)^{-m} |D(\epsilon)|$$

which proves the theorem for the case  $l = m$

The proof for the case  $l < m$  is carried out analogously and relies respectively, on the results of Section 2.7 and formula (4.2.8)

In this manner, for sufficiently large  $N_0$  and  $N$  the value of the determinant  $|D_{ll}(N)|$  is bounded by

$$N^{-l} |D_{ll}(\epsilon)| \leq |D_{ll}(N)| \leq (N - \epsilon)^{-l} |D_l(\epsilon)| \quad (4.2.25)$$

The graphic interpretation of this result is presented in Fig. 19. Curve 1 corresponds to the continuous D optimal designs for  $\sum_{i=1}^n p_i = N$ , curve 2 to the sequential design and curve 3 to the continuous D optimal designs for  $\sum_{i=1}^n p_i = N - \epsilon$ . For small  $N_0$  the values of the corresponding determinants in general need not satisfy (4.2.25).

If the loss following a nonoptimal preliminary experiment were small ( $N \gg \epsilon$ ) then the upper and lower bounds for the determinant  $|D_{ll}(N)|$  will be close together and the characteristics of the sequential designs will not, in practice be different from the characteristics of the statistically D optimal designs.

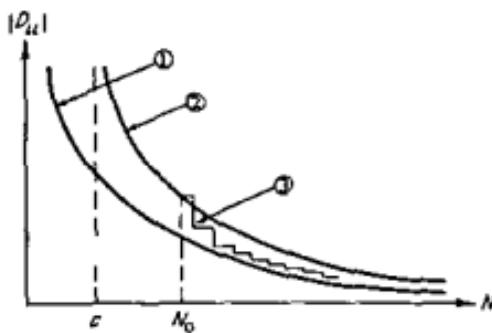


Fig. 19 Graphic interpretation of the properties of sequential designs for large  $N_0$ .

In some cases, particularly for large  $N$  [small  $\lambda(x) d(x, N)$ ], the first equation of system (4.2.8) can be transformed to a simpler equation

$$\lambda(x_N) q'(x_N) D_{ll}^{-1}(N) q(x_N) = \max_x \lambda(x) q'(x) D_{ll}^{-1}(N) q(x),$$

or for an equation equivalent to it [44] (cf. Section 2.7):

$$\lambda(x_N) f'(x_N) \tilde{D}(N) f(x_N) = \max_x \lambda(x) f'(x) \tilde{D}(N) f(x), \quad (4.2.26)$$

where

$$\tilde{D}(N) = \begin{vmatrix} D_{ll}(N) & D_{lk}(N) \\ D_{kl}(N) & D_{kk}(N) - M_{kk}^{-1}(N) \end{vmatrix},$$

$l + k = m$ , and  $M_{kk}(N)$  is the submatrix of the information matrix corresponding to the “unnecessary” parameters. The reduction of the determinant  $|D_{ll}(N)|$  after the  $(N + 1)$ st observation at the point  $x_N$ , satisfying (4.2.26), will be smaller than for the measurement at the corresponding point determined from the first equation of system (4.2.8). The asymptotic properties of the sequential design using (4.2.26) remain as before. This can be shown without difficulty by relying on arguments analogous to those used in Theorems 4.2.1 and 4.2.2.

### 4.3. Sequential Linear-Optimal Designs (Linear Parametrization and Time Constant Efficiency of the Experiment)

Let the assumptions expounded in the discussion to formulas (4.2.2) and (4.2.3) hold. We will consider the design  $\mathcal{E}_1(N)$  preferred to the design  $\mathcal{E}_2(N)$  if

$$L\{D[\mathcal{E}_1(N)]\} < L\{D[\mathcal{E}_2(N)]\},$$

where the functional  $L$  satisfies the requirements (2.9.2)–(2.9.4).

We consider the sequential design, which is carried out according to the following scheme

$$\begin{aligned} \lambda(x_N) L[D(N) f(x_N) f'(x_N) D(N)] &= \max_x \lambda(x) L[D(N) f(x) f'(x) D(N)] \\ D(N+1) &= \left[ I_m - \frac{\lambda(x_N) D(N) f(x_N) f'(x_N)}{1 + \lambda(x_N) d(x_N, N)} \right] D(N). \end{aligned} \quad (4.3.1)$$

If

$$\lim_{N \rightarrow \infty} N\lambda(x_N) d(x_N, N) \neq \infty, \quad (4.3.2)$$

then the following assertion holds

**Theorem 4.3.1.** *The sequential design conducted according to the strategy defined by system (4.3.1) is asymptotically linear-optimal as  $N \rightarrow \infty$*

$$\lim_{N \rightarrow \infty} NL[D(N)] = L[D(\hat{\epsilon})],$$

where  $\hat{\epsilon}$  is a continuous normalized linear-optimal design

*Proof* In system (4.3.1) we pass to normalized designs

$$\begin{aligned} & \lambda(x_N) L[D(\epsilon(N))] f(x_N) f'(x_N) D[\epsilon(N)] \\ &= \max_x \lambda(x) L[D(\epsilon(N))] f(x) f'(x) D[\epsilon(N)] \end{aligned}$$

$$D[\epsilon(N+1)] = (1 - \alpha_N)^{-1} \left[ 1 - \frac{\alpha_N \lambda(x_N) D[\epsilon(N)] f(x_N) f'(x_N)}{1 - \alpha_N + \alpha_N \lambda(x_N) d(x_N, \epsilon(N))} \right] D[\epsilon(N)]$$

where  $\alpha_N = (N+1)^{-1}$

Comparison of the given system and the iterative procedure of Section 2.10 shows that the normalized designs, corresponding to sequential designs  $\epsilon(N)$  defined by system (4.3.1), coincide with the designs  $\epsilon(N)$ , which were constructed at each  $N$ th iteration of the numerical method of constructing linear-optimal designs for  $\alpha_N = (N+1)^{-1}$ . By Theorem 2.10.1 and Remark 2 to this theorem, when the condition

$$\lim_{N \rightarrow \infty} \lambda(x_N) d(x_N, \epsilon_N) = \lim_{N \rightarrow \infty} N\lambda(x_N) d(x_N, \Lambda) < \infty$$

is satisfied the sequence of designs  $\epsilon_N$  converges to the linear optimal design, that is,

$$\lim_{N \rightarrow \infty} L[D(\epsilon_N)] = L[D(\hat{\epsilon})]$$

Passing in the last equation to nonnormalized designs it is not difficult to obtain the validity of the theorem

Let

$$N^{-1} L[D(\hat{\epsilon})] = L[D(N_0)] \quad (4.3.3)$$

where  $\tilde{N} + c = N_0$ ,  $c \geq 0$ , and  $N_0$  is the number of observations allocated to the preliminary experiment, where the estimates of the sought parameters were obtained with covariance matrix  $D(N_0)$ . The following assertion holds.

**Theorem 4.3.2.** *If  $N_0 \gg m$ , then*

$$L[D(N)] \leq (N - c)^{-1} L[D(\hat{\epsilon})], \quad (4.3.4)$$

where  $N \geq N_0$  and  $N - N_0$  is number of measurements allocated to the sequential design.

The proof of this theorem is completely analogous to the proof of Theorem 4.2.2, and is not carried out.

Inequality (4.3.4) cannot be satisfied if condition (4.3.2) is violated. Indeed, in the proof of this inequality the boundedness of the second derivative in  $N$  for the quantities  $(N - c)^{-1} L[D(\hat{\epsilon})]$  and  $L[D(N)]$  is used in an essential way (compare with Theorem 4.2.2). When condition (4.3.2) is violated, the given derivatives can tend to infinity.

For convenience in using system (4.3.1), Table 2 gives the values of the quantity  $L[D(N)f(x)f'(x)D(N)]$  for a number of broadly used criteria of optimality.

The effectiveness of the sequential design at the  $N$ th step can be somewhat increased if the first equation of system (4.3.1) is replaced by the equation

$$\frac{\varphi(x_N, N)}{1 + \lambda(x_N) d(x_N, N)} = \max_x \frac{\varphi(x, N)}{1 + \lambda(x) d(x, N)}, \quad (4.3.5)$$

where

$$\varphi(x, N) = \lambda(x) L[D(N)f(x)f'(x)D(N)].$$

By taking observations at the point  $x_N$ , satisfying (4.3.5), we obtain the largest increase of  $L[D(N)]$  possible for one observation. Changing the first equation of system (4.3.1) requires an increased amount of computation at each step. However, this is unimportant when using fast electronic computing machines. The asymptotic properties (as  $N \rightarrow \infty$ ) of the sequential designs when using (4.3.5) are determined by Theorems 4.3.1 and 4.3.2.

Table 1

Values of the Quantity  $L[D f(x) f(x) D]$   
for a Number of Broadly Used Criteria of Optimality

$L[D]$	$L[D f(x) f(x) D]$
$\text{Tr } D$	$f(x) D^T f(x)$
$\int_Z d(x) dx$	$f(x) D \Omega D f(x) = \int_Z d^2(x) dx$
	$M = \int_Z f(x) f(x) dx$
$LDI$	$[f(x) D]^2$
$I = \ I_1 - I_m\ $	
$d(x_0)$	$d^2(x - x_0)$
$D_{\alpha\alpha}$	$\left[ \sum_{\beta=1}^m f_\beta(x) D_{\alpha\alpha} \right]^2$
$E[(\theta - \hat{\theta}) C(\theta - \hat{\theta})]$	$f(x) D C D f(x) = \text{Tr } q(x) p(x)$
$C = AA$	$q(x) = A D f(x)$

#### 4.4 Sequential Designs for Nonlinear Parametrization

1 Sequential designs are particularly useful for nonlinear parametrization of the response surface. It is possible in this case to construct only locally optimal statistical designs as was shown earlier. The true values of the parameters are not known *a priori* and therefore practical application of locally optimal designs is limited to those cases when the characteristics of the designs  $\mathcal{E}(\theta)$  change slowly in some region  $\Omega$ , which according to the assumptions of the experimenter is known or with large probability contains  $\theta_0$ .

We consider two important cases

- 1 Initial information about the sought parameters  $\theta$  is absent
- 2 There is some initial information concerning the sought parameters  $\theta$

II The value of the matrix  $M$  for any design  $\mathcal{E}(N)$  depends on the

value  $\theta_0$ , which is unknown; therefore, for the construction of any optimal design [here designs minimizing some functional of the covariance matrix  $D(\hat{\theta})$ ], it is necessary to find some region  $\Omega$ , containing the true value of the parameters  $\theta_0$  and in which  $M[\theta, \mathcal{E}(N)]$  varies insignificantly for  $\theta \in \Omega$ . If the initial information about the region of localization  $\theta_0$  is absent, then the experimenter has no alternative but to conduct some “preliminary” experiment for determining this region.

Prior construction of an optimal design of a “preliminary experiment” is impossible if sufficient information about the response surface being studied is not available (only the analytic form of the surface  $\eta(x, \theta)$  and the effectiveness is known). This is because the design  $\mathcal{E}(N)$  achieving

$$\min_{\mathcal{E}(N_0)} \mathcal{L}\{M[\theta, \mathcal{E}(N_0)]\} \quad (4.4.1)$$

usually depends on  $\theta$ . Here  $N_0$  is the number of observations, allocated to the “preliminary” experiment. Therefore the “preliminary” design should be nondegenerate, and give a single-valued estimate of  $\hat{\theta}$ . Usually a design is constructed satisfying this requirement, but this does not present essential difficulty. After conducting a “preliminary” experiment the following alternatives for designing the experiment are possible: Either seek a statistical design  $\mathcal{E}(N - N_0)$ , minimizing the quantity

$$\mathcal{L}\{M[\hat{\theta}(N_0), \mathcal{E}(N_0) + \mathcal{E}(N - N_0)]\},$$

where  $N$  is the number of observations allocated to the entire experiment, and  $\hat{\theta}(N_0)$  are the estimated parameters after  $N_0$  observations, or turn to a sequential design.

The first alternative has the following inadequacies. First, we do not have any guarantee that

$$\min_{\mathcal{E}(N-N_0)} \mathcal{L}\{M[\hat{\theta}(N_0), \mathcal{E}(N_0) + \mathcal{E}(N - N_0)]\}$$

is close to

$$\min_{\mathcal{E}(N-N_0)} \mathcal{L}\{M[\theta_0, \mathcal{E}(N_0) + \mathcal{E}(N - N_0)]\}$$

since  $\hat{\theta}(N_0)$  may be far from  $\theta_0$ . Second, construction of a statistical design considering the initial information about the parameters [ $M \neq 0$ ], as already noted, is a complicated computational

problem. If we restrict ourselves to the construction of the design  $\mathcal{E}(N - N_0)$  realizing

$$\min_{\theta \in V \setminus \{\theta_0\}} \mathcal{L}\{M[\hat{\theta}(N_0), \mathcal{E}(N - N_0)]\}$$

then the sum of the designs  $\mathcal{E}(N - N_0) + \mathcal{E}(N)$  may not be very effective since in constructing the design  $\mathcal{E}(N - N_0)$  no attention should be paid to the behavior of the function  $\eta(x|\theta)$  in those regions of the factor space for which sufficient information has already been obtained [in the realization of the design  $\mathcal{E}(N_0)$ ] and conducting supplementary measurements [in the realization of the design  $\mathcal{E}(N - N_0)$ ] in these regions therefore is not advantageous.

Sequential designs conducted according to the strategy analogous to that presented in Sections 4.2 and 4.3 for the case of linear parametrization is free from the stated inadequacies.

### III Sequential designs in nonlinear parametrization of the response surface $\eta(x|\theta)$ is accomplished in the following manner

1 Some preliminary experiment is conducted according to a nondegenerate design  $\mathcal{E}(N_0)$  using  $N_0$  observations which results in a single valued estimate  $\hat{\theta}(N_0)$  and the approximate value of its covariance matrix  $D[\hat{\theta}(N_0)] \sim M^{-1}[\hat{\theta}(N_0), \mathcal{E}(N_0)]$  (cf. Section 1.4). The number of preliminary measurements  $N_0$  is chosen such that it is sufficient to determine a nondegenerate design  $\mathcal{E}$  the spectrum of which contains a minimal number of points.

2 The coordinates of the point at which it is necessary to take the  $(N+1)$ st observation is determined after each  $N$ th observation from the equation

$$\lambda(x_N) f_N(x_N) \tilde{D}(N) f_N(x_N) = \max_x \lambda(x) f_N(x) \tilde{D}(N) f_N(x) \quad (4.4.2)$$

for  $D$  criteria and truncated  $D$  criteria of comparison of the results of experiments and from the equation

$$\lambda(x_N) L[D(N) f_N(x_N) f_N(x_N) D(N)] = \max_x \lambda(x) L[D(N) f_N(x) f_N(x) D(N)] \quad (4.4.3)$$

for linear criteria of comparison of the results of experiments.

In (4.4.2) and (4.4.3) the following notation was used:

$$f_N'(x) = \left\| \frac{\partial \eta(x, \theta)}{\partial \theta_1}, \frac{\partial \eta(x, \theta)}{\partial \theta_2}, \dots, \frac{\partial \eta(x, \theta)}{\partial \theta_m} \right\|_{\theta=\hat{\theta}_N}, \quad (4.4.4)$$

$$D^{-1}(N) = M[\hat{\theta}(N)] = \sum_{i=1}^N \lambda(x_i) f_N(x_i) f_N'(x_i). \quad (4.4.5)$$

Choices of  $L[D(N)f_N(x)f_N'(x)D(N)]$  may be determined from Table 2 with the corresponding use of (4.4.4) and (4.4.5); the matrix  $\tilde{D}$  is determined in Section 4.2.

Instead of Eqs. (4.4.2) and (4.4.3), it is possible to use equations of the type (4.2.8) and (4.3.5), which by comparison give at each given step a larger increment of information.

3. After each  $(N + 1)$ st observation the matrix  $D(N + 1)$  is computed by formula (4.2.4), with the corresponding change of the function  $f(x)$  to  $f_{N+1}(x)$ .

For a small number of unknown parameters and small  $N$  it is recommended that the covariance  $D(N + 1)$  be computed directly by the method outlined in Section 1.4 using the results of all  $N + 1$  observations. Indeed, formula (4.2.4) permits one to avoid these calculations, which under linear parametrization were introduced under the assumption of independence of the elements of the matrix  $M$  of the estimators  $\hat{\theta}(N)$ . In the current case the elements of the matrix  $M$  depend on the estimates  $\hat{\theta}(N)$ ; therefore the given formula can be used only when  $\hat{\theta}_N$  is not significantly different from  $\hat{\theta}_{N+1}$ .

*Remark.* Sometimes it is useful to conduct several measurements at each step. In this case, asymptotic properties of sequential designs do not change.

The asymptotic optimality of sequential designs conducted according to the strategy 1–3 follows from the consistency of the best quasilinear estimator  $\hat{\theta}$ . Indeed, as  $N \rightarrow \infty$  the estimate  $\hat{\theta}$  approaches  $\theta_0$  with probability one. Therefore, beginning with some  $N(\delta)$ ,  $\delta > 0$ , the matrix  $M[\hat{\theta}(N), \mathcal{E}(N)]$  will be close to  $M[\theta_0, \mathcal{E}(N)]$ , i.e.,

$$\max_{\alpha, \beta} |M_{\alpha\beta}[\hat{\theta}(N), \mathcal{E}(N)] - M_{\alpha\beta}[\theta_0, \mathcal{E}(N)]| < \delta,$$

and it follows that the location of the point  $x_N$  determined by (4.4.2) or (4.4.3) will be close to the point which would be determined as a result of the sequential designs for a given  $\theta_0$ . If  $\theta_0$  is known, then the value of the design for nonlinear parametrization reduces to the

problem of designing experiments for linear parametrization (cf Section 1.4) From this and from Sections 4.2 and 4.3 it follows that the sequential design for known  $\theta_0$  is asymptotically optimal which causes the asymptotic optimality of the design conducted according to steps 1-3 In this manner sequential designs approach locally optimal designs as  $N \rightarrow \infty$

**IV** If there is prior information about the true parameters  $\theta$  (the region of localization of  $\theta_0$  is known) then the necessity of a priming experiment can be dropped and the experimenter can pass to a sequential design conducted according to the strategy defined by (4.2.2) or (4.2.3) It is evident that the asymptotic properties of such a design will be as in the design considered in the preceding section

In order to use Eqs (4.4.2) or (4.4.3) it is necessary to know the matrix  $D(N)$

If the designed experiment is a continuation of previously conducted investigations the results of which are expressed in terms of an estimate and its covariance matrix then when we seek the first optimal point  $x_0$  the preliminary estimate  $\hat{\theta}(N_0)$  is known to us and also its covariance matrix  $D(N_0)$  and the possibility of the application of (4.2.2) or (4.4.3) is obvious

Often prior information about the parameters  $\theta$  is not expressed in terms of the estimators and their covariance matrix As an example of such information we consider the information about the region  $\Omega$  of localization of  $\theta_0$  obtained from theoretical considerations qualitative analysis of the process studied analysis of the results of analogous experiments etc In such cases an elliptical region  $\Omega$  is recommended (the characteristics of which are determined by the matrix  $D$  cf Section 1.3) with the center coinciding (from the point of view preliminary analysis) with the value of the sought parameters with largest probability For example suppose as a result of an analysis of prior information the experimenter can assume that with probability  $p \geq 68\%$  the true value of the parameter  $\theta_0$  is contained in the region  $\Omega$  defined by the inequalities  $\theta_{\alpha 1} \leq \theta_0 \leq \theta_{\alpha 2}$   $\alpha = 1, 2, \dots, m$  Then as  $\hat{\theta}(0)$  it is possible to choose the quantity

$$\hat{\theta}_a(0) = (\hat{\theta}_{\alpha 1} + \hat{\theta}_{\alpha 2})/2$$

and as the covariance matrix

$$D_{aa}(0) = (\hat{\theta}_{\alpha 2} - \hat{\theta}_{\alpha 1})^2/2 \quad D_{ab}(0) = 0 \quad \alpha \neq \beta \quad (4.4.6)$$

Afterward, we rely on the constructive sequential design.

In both cases considered in the previous paragraph, we dealt with the necessity of combining prior information with information obtained as a result of conducting the designed experiment. Combining prior information and new results is usually carried out using the following theorem.

**Theorem 4.4.1.** *Let experiment  $\mathcal{E}_1$  result in the estimate  $\hat{\theta}_1$  with covariance matrix  $D_1$ ; then, after conducting the experiment  $\mathcal{E}_2$  the best linear (quasi linear) estimate of  $\theta$  is*

$$\hat{\theta} = [M(\mathcal{E}_1) + M(\mathcal{E}_2)]^{-1} M(\mathcal{E}_1) \hat{\theta}_1 + [M(\mathcal{E}_2) + M(\mathcal{E}_1)]^{-1} Y(\mathcal{E}_2). \quad (4.4.7)$$

*The covariance matrix of the estimator  $\hat{\theta}$  equals*

$$D(\hat{\theta}) = [M(\mathcal{E}_1) + M(\mathcal{E}_2)]^{-1}. \quad (4.4.8)$$

*Here*

$$M(\mathcal{E}_1) = D^{-1}(\hat{\theta}_1),$$

$$M(\mathcal{E}_2) = \sum_{i=1}^n w_i f(x_i) f'(x_i),$$

$$Y(\mathcal{E}_2) = \sum_{i=1}^n w_i y_i f(x_i),$$

*and  $y_i$  ( $i = 1, 2, \dots, n$ ) are the observations of the experiment  $\mathcal{E}_2$ . The remaining notation is the same as in Section 1.3 or 1.4.*

*Proof.* Let, in the course of experiment  $\mathcal{E}_1$ , the measurement  $\tilde{y}_j$  ( $j = 1, 2, \dots, l$ ) be taken; then by Sections 1.3 or 1.4

$$\hat{\theta} = M^{-1} Y, \quad (4.4.9)$$

where

$$M = \sum_{j=1}^l \tilde{w}_j f(\tilde{x}_j) f'(\tilde{x}_j) + \sum_{i=1}^n w_i f(x_i) f'(x_i) = M(\mathcal{E}_1) + M(\mathcal{E}_2), \quad (4.4.10)$$

and

$$Y = \sum_{j=1}^l \tilde{w}_j \tilde{y}_j f(\tilde{x}_j) + \sum_{i=1}^n w_i y_i f(x_i) = Y(\mathcal{E}_1) + Y(\mathcal{E}_2). \quad (4.4.11)$$

Multiplying the first term in (4.4.11) by  $I_m = M(\mathcal{E}_1)M^{-1}(\mathcal{E}_1)$ , we obtain

$$Y = M(\mathcal{E}_1)\theta_1 + Y(\mathcal{E}_2) \quad (4.4.12)$$

Combining (4.4.9), (4.4.10), and (4.4.12) we obtain

$$\theta = [M(\mathcal{E}_1) + M(\mathcal{E}_2)]^{-1}M(\mathcal{E}_1)\theta_1 + [M(\mathcal{E}_1) + M(\mathcal{E}_2)]^{-1}Y(\mathcal{E}_2),$$

which is what was required to prove

If the design  $\mathcal{E}_2$  is nondegenerate [ $|M(\mathcal{E}_2)| \neq 0$ ], then instead of formula (4.4.7) it is sometimes useful to use its modification

$$\theta = [M(\mathcal{E}_1) + M(\mathcal{E}_2)]^{-1}[M(\mathcal{E}_1)\theta_1 + M(\mathcal{E}_2)\theta_2] \quad (4.4.13)$$

The possibility of using the results of Theorem 4.4.1 in the case where prior information is given in the form of the results of some experiment is obvious. In the other case, the possibility of applying these results does not follow obviously from the proof of the theorem. However, we may take into consideration that for each matrix  $M(0)$  [the inverse matrix  $D(0)$  is defined, for example, by procedure (4.4.6)] it is possible to construct a design "of a fictitious" experiment  $\mathcal{E}_1$  (cf. Theorem 2.1.1), giving the information matrix, equal to  $M(\mathcal{E}_1) = M(0)$ . Repeating the arguments of Theorem 4.4.1 with the "fictitious" experiment  $\mathcal{E}_1$ , it is not difficult to obtain its validity for the case where prior information is obtained from theoretical or qualitative analysis of the process investigated or from other sources.

Assuming that the procedure uses prior information close in spirit to the Bayes approach of constructing estimates (cf., for example, [45, 46]), we can rely on the *a priori* distribution function in the space of unknown parameters. Just as in the Bayes approach, the role of prior information in using (4.4.7) and (4.4.8) decreases with the number of new observations. Indeed, for  $N \rightarrow \infty$ , summing the weights of the new information  $\sum_{i=1}^m w_i \rightarrow \infty$ , expressions (4.4.7) and (4.4.8) take on the form

$$\theta \simeq M(\mathcal{E}_2)Y(\mathcal{E}_2) \quad D(\theta) \simeq M(\mathcal{E}_2)$$

The given approximations say that even in the case of incorrect prior information the results will be close to the true ones for a sufficiently large number of observations. If the prior information were true, then for a small number of measurements we would add essential accuracy

to the results [cf. (4.4.8)]. It is not difficult to see that Theorem 4.4.1 permits the combining also of information obtained from the given experiment and information obtained externally (for example, from new publications) at the time of its realization. The following example is basically (particularly from the chemical viewpoint) the same as the example presented in [41].

EXAMPLE. We consider the chemical reaction of the type



Theoretical analysis of the given reaction shows that it can be described by the model

$$\eta(x, \theta) = \frac{\theta_3 \theta_1 x_1}{1 + \theta_1 x_1 + \theta_2 x_2}, \quad (4.4.15)$$

where  $\eta$  is the speed of the chemical reaction,  $x_1$  is the partial pressure of the sought product  $P$ ,  $x_2$  is the partial pressure of the product  $P_1$ ,  $\theta_2$  is the absorption equilibrium constant for the product  $P_1$ ,  $\theta_3$  is the effective constant of the speed of reaction, and  $\theta_1$  is the absorption equilibrium constant for the reagent  $R$ .

This model is presented for some catalytic reactions of the type (4.4.14), in which the reagent  $R$  is some quaternary or primary alcohol from a long chain, the product  $P_1$  is an olefin, and the product  $P$  is water.

The true values of the parameters  $\theta' = \|\theta_1, \theta_2, \theta_3\|$ , which of course are assumed to be unknown to the experimenter, were found for the catalytically dehydrated *n*-hexyl alcohol at 555°F:

$$\theta_{10} = 2.9, \quad \theta_{20} = 12.2, \quad \theta_{30} = 0.69. \quad (4.4.16)$$

We assume that the aim of the experiment is to determine as accurately as possible the parameters  $\theta_1, \theta_2, \theta_3$  and as the measure of accuracy use the determinant of the information matrix. In what follows, we will consider that the observations are possible in the region

$$0 \leq x_1 \leq 3, \quad 0 \leq x_2 \leq 3.$$

The results of the observations at the point  $x' = \|x_1, x_2\|$ , belonging to the given region, will be obtained using a table of random numbers. It is assumed that  $y$  has a normal distribution around the mean  $\eta(x, \theta_0)$  with a constant variance equal to  $b^2 = 0.01$ .

The preliminary estimates of the parameters  $\theta$  were obtained from an experiment conducted at four points (cf. Table 3) and are equal to

$$\theta_1(4) = 10.39 \quad \theta_2(4) = 48.83 \quad \theta_3(4) = 0.74$$

Table 3  
Preliminary Estimates of the Parameters  $\theta$

Experiment number	$x_1$	$x_2$	$y$
1	1	1	0.126
2	2	1	0.219
3	1	2	0.076
4	2	2	0.126

Assuming that the estimate  $\theta(4)$  is not very far from the true value (in the contrary case the first step of the sequential design will not be very effective) we use the method developed above. We compute the function  $f_{t(4)}(x)$

$$f_{t(4)}(x) = \frac{\partial}{\partial \theta_a} \left. \frac{\theta_3 \theta_1 x_1}{1 + \theta_1 x_1 + \theta_2 x_2} \right|_{\theta = \theta(4)} \quad (4.4.17)$$

or, more explicitly

$$f_{1(4)}(x) = \frac{0.74 x_1 + 36.1 x_1 x_2}{(1 + 10.39 x_1 + 48.83 x_2)^2}$$

$$f_{2(4)}(x) = \frac{7.69 x_1 x_2}{(1 + 10.39 x_1 + 48.83 x_2)^2}$$

$$f_{3(4)}(x) = \frac{10.39 x_1}{(1 + 10.39 x_1 + 48.83 x_2)}$$

After four observations the information matrix with accuracy up to a constant multiplier will have the form

$$M(4) = \sum_{i=1}^4 f_{t(4)}(x_i) f_{t(4)}(x_i) \quad (4.4.18)$$

where the values of the coordinates of the points  $x_i$  ( $i = 1, 2, 3, 4$ ) are given. In constructing (4.4.18) it was assumed that all observations

are equally distributed. Using the values  $x_i$  ( $i = 1, 2, 3, 4$ ) and inverting the matrix  $M(4)$ , we construct the quantity

$$d(x, 4) = f'_{(4)}(x)D(4)f_{(4)}(x),$$

and solve the equation

$$d(x_5, 4) = \max_x d(x, 4). \quad (4.4.19)$$

We note that the quantity  $d(x, 4)$  is approximately equal to the variance of the estimate of the response surface ( $d(x, 4) \approx d[\hat{\eta}(x, 4)]$ ), in which case the accuracy of the approximation is determined by the smoothness of  $\eta(x, \theta)$  as a function of  $\theta$  and the closeness of  $\hat{\theta}(4)$  to  $\theta_0$  (cf. Section 1.4). The solution of (4.4.19) is the point (and in particular at this point it is necessary to take the fifth observation) with coordinates

$$x_{1(5)} = 0.1, \quad x_{2(5)} = 0.0.$$

The search for the solution is conducted by a method of nets. [The values  $d(x, 4)$  are computed at each of  $31 \times 31$  points  $x_j$ .]

The experiment "conducted" at the point  $x_5$  gave the result  $y_5 = 0.186$ . Using this, we obtain

$$\hat{\theta}'(5) = \| 3.11, 15.19, 0.79 \|.$$

The derivatives in (4.4.18) are computed for  $\theta = \hat{\theta}(5)$ , and we set up the matrix  $M(5)$ . Then we seek the maximum in  $x$  of the quantity  $d(x, 5)$  (the maximum is attained for  $x_6' = \| 3.0, 0.0 \|$ ). In Table 4 the values of the function  $d(x, 5)/d(x_6, 5)$  are presented for a  $7 \times 7$  grid. After this we take a measurement at the point  $x_6$  and find the estimate  $\hat{\theta}(6)$ , etc. Then, in a completely analogous manner, the seventh obser-

**Table 4**  
The Values of the Function  $d(x, 5)/d(x_6, 5)$  for a  $7 \times 7$  Grid

$x_2$	0	0.5	1.0	1.5	2.0	2.5	3.0
3.0	0.0015	0.0015	0.0019	0.0019	0.0019	0.0019	0.0019
2.5	0.0015	0.0015	0.0019	0.0019	0.0022	0.0019	0.0019
2.0	0.0015	0.0019	0.0019	0.0019	0.0019	0.0019	0.0022
1.5	0.0015	0.0019	0.0019	0.0019	0.0019	0.0022	0.0030
1.0	0.0015	0.0019	0.0019	0.0019	0.0030	0.0055	0.0101
.5	0.0015	0.0019	0.0026	0.0082	0.0212	0.0416	0.0676
0	0.0015	0.0145	0.429	0.645	0.800	0.914	1.0

vation is planned. The results of the design and the experiment are presented in Table 5.

Table 5

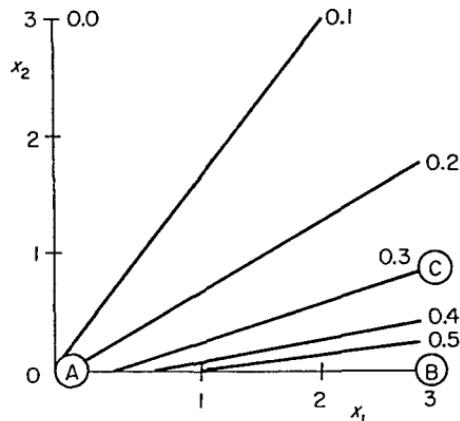
## The Results of the Design and the Experiment

Observation number	$x_1$	$x_2$	$y$	$\theta_1$	$\theta_2$	$\theta_3$
1	10	10	0.126			
2	2.0	10	0.129			
3	10	20	0.076			
4	20	20	0.126	10.39	48.83	0.74
5	0.1	0.0	0.186	3.11	15.19	0.79
6	3.0	0.0	0.606	3.96	15.32	0.66
7	0.2	0.0	0.268	3.61	14.00	0.66
8	3.0	0.0	0.614	3.56	13.96	0.67
9	0.3	0.0	0.318	3.32	13.04	0.67
10	3.0	0.8	0.298	3.33	13.48	0.67
11	3.0	0.0	0.509	3.74	13.71	0.63
12	0.2	0.0	0.247	3.58	13.15	0.63
13	3.0	0.8	0.319	3.57	12.77	0.63

The modeling of the experiment was stopped as soon as the dispersion of the parameter estimates became insignificant and the value of the estimate became close to the theoretical value of the parameters  $\theta_1, \theta_2, \theta_3$ . In practice, such an experiment is usually stopped if the determinant  $|D(N)|$  or the risk function (cf. Section 1.9) becomes less than some number given beforehand.

We consider the results of the design carried out. It is not difficult to see that the values  $x_i$  ( $i = 5, 6, \dots, 13$ ) are concentrated close to three points A, B, C (cf. Fig. 20).

For clarity the lines of constant values of  $\eta(x, \theta_0)$  are drawn in this figure. Corresponding to Theorem 4.2.1 and the results of the current section these points must correspond to a locally  $D$ -optimal design for zero initial information ( $M = 0$ ). The number of observations  $N_1, N_2, N_3$  ( $N_1 + N_2 + N_3 = N$ ) which are needed at each point of a locally  $D$ -optimal design must be proportional to the frequency of occurrence of these points in the sequential design (cf. the results of Section 2.5). For a precise determination of the characteristics of the optimal design it is necessary to consider about 100 iterations of the



**Fig. 20 [41].** A contour diagram for  $\eta_0$  as a function of  $x_1$  and  $x_2$  for the model (4.4.15). The circles indicate the position of points of observation obtained by methods of sequential design.

type (4.4.17)–(4.4.19). The following design is the locally  $D$ -optimal design for  $\theta_0$ , defined by (4.4.16):

$$\dot{\epsilon}(\theta_0) = \left\{ \begin{array}{l} x_1' = \|0.2, 0.0\|; x_2' = \|3.0, 0.0\|; x_3' = \|3.0, 1.0\| \\ p_1 = p_2 = p_3 = \frac{1}{3} \end{array} \right\}.$$

In Table 6 the characteristics of several designs are presented. As was to be expected, Design 3 is the best. This design, however, cannot be constructed *a priori* ( $\theta_0$  in real problems is not known). After it follows Design 2, then 1. The design constructed at the vertex of the region (cf. Fig. 20) is degenerate.

**Table 6**  
The Characteristics of Several Designs

No.	Characteristics of the designs	Ratio $ D_s  /  D[\dot{\epsilon}(\theta_0, 12)] $
1	$x_1 = \ 1.0, 1.0\ ; x_2 = \ 1.0, 2.0\ $ $x_3 = \ 2.0, 1.0\ ; x_4 = \ 2.0, 1.0\ $ $N_1 = N_2 = N_3 = N_4 = 3$	$\sim 0.5(10^6)$
2	First 12 points from Table 5 $N_i = 1 (i = 1, 2, \dots, 12)$	$\sim 1.5$
3	$x_1 = \ 0.1, 0.0\ ; x_2 = \ 3.0, 0.0\ ; x_3 = \ 3.0, 0.8\ $ $N_1 = N_2 = N_3 = 4$	$\sim 1.0$

In this example, for the search of the maximal value of  $d(x, N)$ , the method of nets was applied. This method is somewhat cumbersome from the computational point of view, however, it gives the experimenter information which is difficult or impossible to obtain by other methods, for example, gradient methods. Indeed, having the data of the type of Table 4, the experimenter can easily see how accurately he must fix each of the control variables.

Thus, for example, from Table 4 it follows that particular attention must be paid to fixing the coordinate  $x_2$ , the partial pressure of the reagent  $R$ , and significantly less attention can be paid to fixing coordinate  $x_1$ , the partial pressure of the product  $P$ .

The sequential design of the preceding experiment was conducted according to the strategy which permitted the maximization of the determinant of the information matrix. If the experimental situation is such that a more suitable measure of accuracy of the experiment is some other quantity, for example, the mean dispersion of the estimates of the parameters, then at each stage of the sequential design one must seek correspondingly

$$\max_x f_N(x) D^*(N) f_N(x)$$

The results of the corresponding sequential design of the experiment are presented in Table 7. In this table the characteristics of the local

Table 7  
The Results of the Corresponding Sequential Design of the Experiment

Number	Characteristics of the designs	Ratio $\text{Tr } D_p / \text{Tr } D[\delta(\theta_0, 12)]$
1	$x_1 = [10 10] \quad x_2 = [10 20]$ $x_3 = [20 20] \quad x_4 = [20 10]$ $N_1 = N_2 = N_3 = N_4 = 3$	$\sim 2.3(10^4)$
2	$x_1 = [10 10] \quad x_2 = [10 20] \quad x_3 = [20 20]$ $x_4 = [20 10] \quad x_5 = [03 00] \quad x_6 = [30 06]$ $x_7 = [03 00] \quad x_8 = [30 06] \quad x_9 = [03 00]$ $x_{10} = [30 075] \quad x_{11} = [03 00] \quad x_{12} = [30 075]$	$\sim 1.5$
3	$x_1 = [03 00] \quad x_2 = [30 00]$ $x_3 = [30 06] \quad x_4 = [30 075]$ $N_1 = 6 \quad N_2 = N_3 = 1 \quad N_4 = 4$	1.0

*A*-optimal design are presented as well as the design, all measurements of which are concentrated at four points chosen beforehand.

In conclusion, we note that the example illustrates also the fundamental distinction between the problem, the aim of which is to estimate the response surface  $\eta(x, \theta)$  and the problem, the aim of which is to estimate the parameters  $\theta$ . For example, after the fourth observation, the estimate for  $\eta(x, \theta)$  in the region where the observations were taken is found to be in good agreement with the experiment. At the same time the estimates  $\hat{\theta}_1(4)$  and  $\hat{\theta}_2(4)$  are quite different from the true values of the corresponding parameters.

This fact characterizes the compensatory nature of the errors of the estimates  $\hat{\theta}_1(4)$  and  $\hat{\theta}_2(4)$ . Although both estimates are much larger than the corresponding true values, the errors are such that if the estimates are substituted in place of  $\theta_1$  and  $\theta_2$ , the results in terms of the estimate of the quantity  $\eta(x, \theta)$  in the region where the observations were conducted are in good agreement with the true values. Such a correlation among the estimates is characteristic for models of the considered type. In particular, such correlations are characteristic of the estimates of the parameters for the investigated catalytic reaction kinetics, since they often have a form analogous to (4.4.15).

#### 4.5. Designs When the Efficiency Function of the Experiment Is Unknown

I. In many experimental investigations the efficiency function  $\lambda(x)$  of the experiment is unknown or essentially unknown. Statistical designs in this case are impossible (recall that for a statistical design it is necessary to know the efficiency function of the experiment with accuracy up to a constant multiplier). In this case we investigate the possibility of sequentially designing experiments.

We will assume that the function  $\lambda(x)$  is sufficiently smooth in the region  $X$ , where the observations are taken, or its analytic form  $\lambda(x, \omega)$  is known as a function of some unknown parameters  $\omega' = [\omega_1, \omega_2, \dots, \omega_k]$ . In what follows we will consider only the last case. The first case is reduced to the second by means of an approximation of  $\lambda(x)$  by some polynomial  $\sum_{a=1}^k \omega_a \varphi_a(x)$  [this is possible in view of the smoothness of  $\lambda(x)$ ].

We note that the efficiency function of the experiment by definition is always nonnegative. Therefore, for approximating it by some poly-

nomial, the estimates of the coefficients  $\omega$  must be sought only among the set for which the inequality  $\sum_{a=1}^k \omega_a \varphi_a(x) \geq 0$  is satisfied. Usually this requirement calls for additional computational difficulties. In order that they be avoided it is recommended that  $\ln \lambda(x)$  be approximated instead of  $\lambda(x)$ . In this case we do not have the additional bound for the approximating polynomial since the logarithm of the efficiency function can take on any value between  $-\infty$  and  $+\infty$ .

**II.** We consider the following strategy for conducting an experiment

1 A design  $\mathcal{E}(N_0)$  is constructed which permits simultaneous determination of estimates of the functions  $\lambda(x, \omega)$  and  $\eta(x, \theta)$ . At each point of the design  $x_i$  [ $i = 1, 2, \dots, n \geq \max(k, m)$ ] several observations  $n_i$  ( $\sum_{i=1}^n n_i = N_0$ ) must be taken in order to compute an estimate of the dispersion  $d_i$  of the observed quantities  $y_i$ . The estimate of the dispersion is computed according to the formula

$$d_i = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1} \quad \bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij} \quad (4.5.1)$$

As an estimate of the parameters  $\omega$  it is possible to use the quantities  $\hat{\omega}(N_0)$  corresponding to

$$\min_{\omega} \sum_{i=1}^n [d_i^{-1} - \lambda(x_i, \omega)]^2 \quad (4.5.2)$$

It is possible to verify that the estimate  $\hat{\omega}$  is consistent (cf. Section 1.4). For a preliminary estimate of the parameters  $\theta$  the quantities  $\hat{\theta}(N_0)$  are chosen corresponding to

$$\min_{\theta} \sum_{i=1}^n w_i [y_i - \eta(x_i, \theta)]^2 \quad (4.5.3)$$

where  $w_i = d_i^{-1} n_i$ . If the values  $n_i$  ( $i = 1, 2, \dots, n$ ) are not too small, the estimates  $\hat{\theta}(N_0)$  will be close, according to their characteristics to the best linear (quasi linear) estimates. Computation of  $\hat{\theta}(N_0)$  can be carried out according to formulas of Sections 1.3 and 1.4 with  $w_i$  replaced by  $d_i^{-1} n_i$ . The dispersion matrix of the estimates  $\hat{\theta}(N_0)$  is determined by the approximation formula

$$D[\hat{\theta}(N_0)] \simeq \left[ \sum_{i=1}^n w_i f(x_i) f'(x_i) \right]^{-1} \quad (4.5.4)$$

where

$$f_\alpha(x) = \partial\eta(x, \theta)/\partial\theta_\alpha|_{\theta=\hat{\theta}(N_0)}.$$

The degree of approximation is determined by the accuracy of the estimates  $\hat{d}_i$  ( $i = 1, 2, \dots, n$ ) and (for the case of nonlinear dependence of the response surface on the parameters) the closeness of  $\hat{\theta}(N_0)$  to the true values  $\theta_0$  (cf. Section 1.4).

2. After the preliminary estimates of the efficiency function of the experiment  $\hat{\lambda}_0(x) = \lambda[x, \hat{\omega}(N_0)]$  are found, the point  $x_0$  is sought from Eqs. (4.4.2) or (4.4.3) (depending on the choice of criteria of optimality). In this case, in the given equations it is necessary to replace  $\lambda(x)$  by  $\hat{\lambda}(x)$ . At the point  $x_0$ ,  $\hat{n}$  observations are taken.

3. The quantities  $\hat{d}_1, y_1$  are found and the estimates  $\hat{\lambda}_1(x) = \lambda[x, \hat{\omega}(N_0 + \hat{n})]$  and  $\hat{\theta}(N_0 + \hat{n})$  are then computed.

4. From Eqs. (4.4.2) or (4.4.3) with  $\lambda(x) = \hat{\lambda}_1(x)$  the point  $x_1$  is determined,  $n_1$  observations are taken at this point, and so on.

The procedure defined by 2–4, is continued as long as the experimenter does not attain the required accuracy of determination of the parameters in the sense of his chosen criteria.

Since  $\hat{\omega}$  and  $\hat{\theta}$  are consistent estimates, it is not difficult to obtain asymptotic optimality of the experimental strategy presented by repeating considerations of the preceding section.

**III.** Taking several observations at each point  $x_s$  is necessary for a determination of the quantities  $\hat{d}_s$ . The allocation of an extremely large number of observations at one point, however, has a negative effect on the optimal properties of sequential design (cf. the proofs of the theorems from Sections 4.2 and 4.3). Therefore, for choosing the quantity of observations at each point  $x_s$  it is useful to be guided by the fact that the optimal properties, of the methods considered in the present chapter on sequential designs, are preserved for  $\hat{n} \ll N$ . As soon as the function  $\lambda(x)$  is defined sufficiently accurately, for example, beginning with some  $s$  the inequality

$$\max_x |\hat{\lambda}_s(x) - \hat{\lambda}_{s-1}(x)| \leq \delta$$

is satisfied, where  $\delta$  is an “accuracy” given beforehand, it is possible to take one observation at each point  $x_s$ . When this is done, we set the weights  $w_i = \hat{\lambda}_s(x_i)$ . Sometimes the efficiency function  $\lambda(x)$  depends on the unknown parameters themselves, and also on the

response surface  $\eta(x, \theta)$ . In these cases it is not necessary to find the minimum of (4.5.2), in the remaining design the procedure remains the same.

#### 4.6 Design in the Presence of Errors in the Determination of the Control Variables

We will assume that the coordinates of the point  $x$ , at which the observations are taken, are given with some error (for more details cf. Part I of Section 1.6). The dispersion matrix of the estimator  $\hat{\theta}$ , constructed by the methods presented in Section 1.6, is equal to

$$D(\hat{\theta}) \simeq \left[ \sum_{i=1}^n w_i(\hat{\theta}) \varphi(\hat{\theta}_i, x_i) \varphi(\hat{\theta}, x_i) \right]^{-1}, \quad (4.6.1)$$

where the weights  $w_i$  and  $\varphi(\hat{\theta}, x_i)$  are determined in the same way as in Section 1.6.

If the criterion of optimality of the experiment is a criterion determined by elements of the matrix  $D(\theta_0)$ , then it is possible to construct only locally optimal designs since the elements of the dispersion matrix of the estimated parameters depend on the true values of the true parameters. Sequential design of experiments, in the presence of errors in the determination of the coordinates of the points where observation are taken, can be carried out according to the following strategy:

1 A nondegenerate design  $\mathcal{E}(N_0)$  is constructed according to the methods presented in Section 1.6 and the estimates  $\hat{\theta}(N_0)$  are determined.

2 The point  $x_0$  is found, which is a solution to Eqs. (4.4.2) or (4.4.3), where

$$\lambda(x) = \{\tilde{\lambda}^{-1}(x) + \nabla_1^2[\hat{\theta}(N_0) \cdot x] d_2(x)\}^{-1} \quad (4.6.2)$$

and  $\tilde{\lambda}(x)$  is the efficiency function of the experiment in the absence of errors in  $x$  [ $\tilde{\lambda}(x_i) = b_i^{-2}$ ]. The matrix  $D(N_0)$  is determined from (4.6.1).

3 At the point  $x_0$  a measurement is taken (or several measurements). Again the estimate  $\hat{\theta}(N_0 + 1)$  is computed.

4 Part 2 is repeated with the estimate  $\hat{\theta}(N_0 + 1)$ , and the next observation is taken at the point  $x_1$ . This is continued as long as the

accuracy of determination of the unknown parameters (in the sense of the chosen criteria for comparing the results of experiments) does not attain a given value.

Since the estimates  $\hat{\theta}$ , obtained by the method presented in Section 1.6 are consistent it is not difficult to obtain the asymptotic optimality of the design of the experiment by repeating the arguments of Section 4.4.

We note that for sequential designs the measurements will be allocated basically in those regions where the influence of errors in the determination of the control variables is small. Indeed, the influence of errors in  $x$  basically determine the values of the first derivatives  $\nabla_1(\theta, x)$ . But from (4.6.2) and (4.4.2) or (4.4.3) it follows that the measurements will be allocated where the values  $[\nabla^2(\theta, x) d_2(x)]$  are not very great (the influence of the errors in  $x$  is insignificant).

#### 4.7. Construction of Optimal Designs When the Experimental Conditions Vary in Time

I. As noted in Section 4.1, in long-duration experiments the efficiency function of the experiment can vary in time for a number of reasons. In this case the optimal design must determine not only the coordinates of the points at which it is necessary to take observations, but also when at the given points of the design it is necessary to take measurements.

If the behavior of the efficiency function  $\lambda(x, t)$  in time is known beforehand, then the construction of exact or discrete statistically optimal designs (i.e., considering the discreteness of the allocations, cf. Chapter 3) is possible in principle, although it is a very complicated computational problem.

For linear parametrization it is helpful in solving such problems to use the methods of constructing discrete designs presented in Sections 3.2 and 3.3. If in the course of taking each observation  $\lambda(x, t)$  does not change, but the observation is conducted at a given moment in time  $t_i$  ( $i = 1, 2, \dots, N$ ), then an iterative procedure analogous to that considered in Sections 3.2 and 3.3 is constructed in the following manner.

1. Let there be a design  $\mathcal{C}_0(N)$  with spectrum  $x_1, x_2, \dots, x_N$ . For each  $i$ th observation taken at the moment  $t_i$  the quantity  $\Delta_0(x_i, x, t_i)$  is constructed. The form of  $\Delta_0(x_i, x, t_i)$  corresponds to the criteria of

optimality used (cf. either Sections 3.2 or 3.3). In the formulas introduced in the said sections it is necessary to replace  $\lambda(x_i)$  and  $\lambda(x)$  by  $\lambda(x_i, t_i)$  and  $\lambda(x, t)$ .

2 The following is computed

$$\max_i \max_x D_i(v_i, x, t_i) \quad (i = 1, 2, \dots, N) \quad (4.7.1)$$

3 Let the maximum be obtained for  $x = \hat{x}$  and  $i = j_0$ . The design  $\mathcal{E}_1(N)$  is constructed which has the spectrum  $x_1, x_2, \dots, x_{j_0-1}, x_{j_0+1}, \dots, x_N, \hat{x}$ . For this design the dispersion matrix  $D[\mathcal{E}_1(N)]$  is found.

4 Operations 1-3 are repeated with the design  $\mathcal{E}_1(N)$ , then with the design  $\mathcal{E}_2(N)$ , and so on as long as the following inequality is not satisfied

$$\max_i \max_x D_i(x_i, x, t_i) < \delta, \quad (4.7.2)$$

where  $\delta$  is an accuracy given beforehand.

It is easy to verify that the procedure presented converges to a design  $\mathcal{E}(N)$  which is better than the initial one but which may not be optimal. In connection with this, it is recommended that Steps 1-4 be carried out with several initial approximations.

II. Let

$$\lambda(v, t) = \lambda(v) l(t), \quad 0 \leq t \leq t_0, \quad (4.7.3)$$

and let the continuous normalized design

$$\hat{\epsilon} = \left\{ \frac{x_1}{p_1}, \frac{x_2}{p_2}, \dots, \frac{x_n}{p_n} \right\} \quad (4.7.4)$$

be optimal ( $D$ - or linear-optimal) for the efficiency function of the experiment equal to  $\lambda(x)$ .

We consider the continuous design  $\hat{\epsilon}(t)$ , at which the distribution of the allocation, in the factor space and in time, is accomplished according to the following method.

$$\hat{\epsilon}(t) = \left\{ \frac{x_1}{p_1(t)}, \frac{x_2}{p_2(t)}, \dots, \frac{x_n}{p_n(t)} \right\}, \quad (4.7.5)$$

where  $\hat{l}(t) = cl(t)$ , the normalizing multiplier, is chosen such that the sum of the allocation is equal to  $T$ .

**Theorem 4.7.1.** Let the design  $\hat{\epsilon}$  for a constant efficiency function of the experiment be D- or linear-optimal; then the design  $\hat{\mathcal{E}}(t)$  is D- or linear-optimal for the efficiency function of the experiment defined by (4.7.3).

*Proof.* We carry out the proof for D-optimality. For the remaining criteria (truncated D-criteria, linear criteria) the proof is carried out in an identical manner.

The information matrix for the design  $\hat{\mathcal{E}}(t)$  has the form

$$M[\hat{\mathcal{E}}(t_0)] = \int_0^{t_0} \sum_{i=1}^n p_i \lambda(x_i) \tilde{l}(\tau) f(x_i) f'(x_i) d\tau,$$

or

$$\begin{aligned} M[\hat{\mathcal{E}}(t_0)] &= \int_0^{t_0} \tilde{l}(\tau) d\tau \sum_{i=1}^n p_i \lambda(x_i) f(x_i) f'(x_i) \\ &= M(\hat{\epsilon}) \int_0^{t_0} \tilde{l}(\tau) d\tau = M(\hat{\epsilon}) T. \end{aligned} \quad (4.7.6)$$

On the other hand, for any other design the information matrix has the form

$$M[\mathcal{E}(t_0)] = \int_0^{t_0} \tilde{l}(\tau) \int_X p(x, \tau) \lambda(x) f(x) f'(x) dx d\tau, \quad (4.7.7)$$

where  $p(x, t)$  is the normalized distribution of allocation for the design  $\mathcal{E}(t)$  at the time  $t$ . Dividing (4.7.7) by  $\int_0^{t_0} \tilde{l}(\tau) d\tau = T$ , we can treat the matrix

$$M(\epsilon) = M[\mathcal{E}(t)] \cdot T^{-1}$$

as a linear combination of information matrices

$$M[\epsilon(\tau)] = \int_X p(x, \tau) \lambda(x) f(x) f'(x) dx,$$

corresponding to the designs with the distribution of allocation  $p(x, \tau)$  and the efficiency function  $\lambda(x)$  which is constant in time. The corresponding design

$$\epsilon = T^{-1} \mathcal{E}(t)$$

is a linear combination of these designs. But any linear combination

of designs with efficiency function constant in time gives the design  $\epsilon$  for which [cf Theorem 2.2.2]

$$|M(\epsilon)| \leq |M(\epsilon)| \quad (4.7.8)$$

Since [cf (4.7.6) and (4.7.7)]

$$|M[\mathcal{E}(t)]| = T^m |M(\epsilon)| \quad (4.7.9)$$

and

$$|M[\mathcal{E}(t_0)]| = T^m |M(\epsilon)| \quad (4.7.10)$$

then from (4.7.8)–(4.7.10) it follows that

$$|M[\mathcal{E}(t_0)]| \leq |M[\mathcal{E}(t_0)]|$$

which is what was required to be proved

The optimality of designs of the type (4.7.5) was first noted in [41] for  $D$  optimal designs.

From Theorem 4.7.1 it follows that for  $D$  and linear criteria of comparison of experiments [more precisely for any criteria for which (3.1.3) holds] the continuous design  $\mathcal{E}(t)$  constructed using the continuous normalized design  $\epsilon$  (corresponding to  $D$  or linear optimality) is optimal. Analogous results can be obtained also for locally optimal designs.

If the number of observations  $N$  used in the experiment is large and  $I(t)$  varies little in time then alternating observations at the points  $x_1, x_2, \dots, x_n$  with frequency  $p_1, p_2, \dots, p_n$  we obtain a design close to optimal.

III It follows immediately from the proof of Theorems 4.2.1 or 4.3.1 that the sequential design (for efficiency functions constant in time and with large  $\Lambda$ ) is carried out by distributing the observations at points (actually close to points)  $x_1, x_2, \dots, x_n$  with frequency  $p_1, p_2, \dots, p_n$ . It is evident that if  $I(t)$  varies little in the time  $\Delta t$  in the course of which the frequency of taking observations at the points  $x_1, x_2, \dots, x_n$  will be sufficiently well approximated by the quantities  $p_1, p_2, \dots, p_n$  then the sequential design will be optimal. For a good approximation the quantities  $p_1, p_2, \dots, p_n$  must satisfy the requirement

$$\Delta t \ll \Delta T \quad (4.7.11)$$

where  $\Delta t$  is the time necessary to take one observation

The representation of the efficiency function in the form (4.7.3) describes the majority of real experimental situations fairly well. Indeed, the efficiency function, as a rule, changes either as a result of improvement of the measuring apparatus or as a result of changing the amount of the substances participating in the reaction under study. There can be other reasons for the efficiency function changing in time, which, as the two cited above, usually satisfy natural requirements of uniformly worsening or bettering the experimental conditions in the entire region  $X$  where the observations are possible.

It is also to be pointed out that the function  $l(t)$  very often has the form presented in Fig. 21. The points where  $l(t)$  experiences jumps correspond, for example, to a change of the measuring apparatus to a more accurate one. It is unlikely that in the course of one experiment the apparatus changes very often. Therefore condition (4.7.11) is a sufficiently weak restriction.

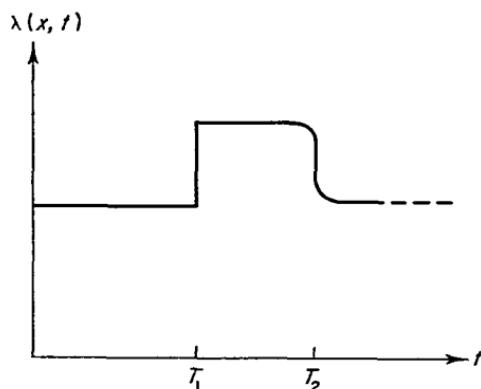


Fig. 21. The time  $T_1$  may correspond, for example, to a change in apparatus, the time  $T_2$  to an increase in external assistance.

We will assume that the function  $\lambda(x, t)$  changes in time and that jumps analogous to those presented in Fig. 21 are rare. The number of jumps can be reduced for increasing  $x$ . The time of the jumps  $T_j$  (and sometimes the existence of such jumps) usually cannot be predicted beforehand. In such situations the optimal strategy is the following. In the interval from 0 to  $T_1$  the experimenter takes measurements according to the optimal design for  $\lambda(x, t \ll T_1)$ .

After realization of the given design, the dispersion matrix of the estimators of the unknown parameters becomes equal to  $D(T_1)$ . From the time  $T_1$  to the time  $T_2$  the measurements are taken according to the optimal design for  $\lambda(x, T_1 \leq t \leq T_2)$  and the initial conditions  $D[\theta(T_1)]$ . Furthermore, the optimal design is constructed for

$\lambda(x, T_2 \leq t \leq T_3)$  and the initial conditions  $D[\theta(T_2)]$ , and so on. Let the interval between jumps be sufficiently large. Then, carrying out arguments analogous to the arguments of Part III of Section 4.4, it is not difficult to show that, using in each interval the methods developed above for sequential design, the experimenter will obtain

$$D_s(\theta, T_j) \simeq D[\theta(T_j)]$$

The index  $s$  indicates that the given quantity is obtained by realization of the sequential design.

# 5

## Design of Experiments in the Case of Simultaneous Observation of Several Random Quantities

### 5.1. Basic Properties of the Information Matrix

I. We generalize the results of Chapters 2–4 to the case when, at one and the same point of the factor space, simultaneous measurements of several quantities  $y' = \|y_1, y_2, \dots, y_k\|$  (in the general case correlated among themselves) [47] are possible.

Let the analytic form of the function

$$\eta'(x, \theta) = \|\eta_1(x, \theta), \eta_2(x, \theta), \dots, \eta_k(x, \theta)\| \quad (5.1.1)$$

be known. Also let the covariance matrix of the quantity  $y$  be known at each point  $x$ :

$$D(y | x) = w^{-1}(x). \quad (5.1.2)$$

The matrix  $w(x)$  will be called the weight matrix. Together with the weight matrix  $w(x)$  it will be useful for us to use the matrix of efficiency of the experiment:

$$\lambda(x) = w(x)/r, \quad (5.1.3)$$

where  $r$  is the number of observations taken at the point  $x$ .

In Section 1.7 it was shown that after  $n$  observations  $y_1, y_2, \dots, y_n$  the best linear estimate for the parameters  $\theta$  in the case of linear parametrization of the surface  $\eta(x, \theta)$  is the quantity

$$\hat{\theta} = M^{-1}Y \quad (5.1.4)$$

Their covariance matrix is the inverse of the information matrix

$$D(\hat{\theta}) = M^{-1} \quad (5.1.5)$$

The matrices  $M$  and  $Y$  are defined in Section 1.7.

The best quasi-linear estimators and their accuracy are defined corresponding to formulas (5.1.4) and (5.1.5) by replacing in them (cf. Section 1.7)

$$f_{j\alpha}(x) \quad \text{by} \quad \partial \eta_j(x, \theta) / \partial \theta_\alpha |_{\theta=\theta_0},$$

where  $j = 1, 2, \dots, k$  and  $\alpha = 1, 2, \dots, m$ .

II. Just as earlier, we indicate the normalized continuous designs by  $\epsilon$ , and the nonnormalized by  $\mathcal{E}(N)$ . We investigate the properties of the information matrix  $M(\epsilon)$ .

**Theorem 5.1.1.** *The matrix  $M(\epsilon)$  has the following properties*

- (1) *For any design  $\epsilon$  the matrix  $M(\epsilon)$  is positive semidefinite,*
- (2)  *$|M(\epsilon)| = 0$  for any design  $\epsilon$  containing less than  $m/k$  points,*
- (3) *The set of matrices  $M(\epsilon)$ , corresponding to arbitrary designs  $\epsilon$  defined on  $X$ , is convex and closed,*
- (4) *For any design  $\epsilon_1$  with given matrix  $M(\epsilon_1)$  a design  $\epsilon_2$  can always be found, the spectrum of which contains no more than  $m(m+1)/2 + 1$  points*

*Proof* (1) The matrices  $F(x_i) w(x_i) F(x_i)^T$  are positive semidefinite [for the definition of the matrix  $F(x)$ , cf. Section 1.7], since

$$l^T F(x_i) w(x_i) F(x_i)^T l = q_i^T w(x_i) q_i \geq 0,$$

where  $l$  is an  $m \times 1$  vector,  $q_i = F'(x_i)l$  is a  $k \times 1$  vector, and the matrix  $w(x_i)$  is positive definite by definition. It follows that the matrix  $M(\epsilon)$  appearing as the sum of these matrices is also positive semidefinite.

- (2) The matrix  $M(\epsilon)$  is the sum of  $n$  matrices each of which has

rank no more than  $k$ . If  $n < m/k$ , then the rank of  $M(\epsilon)$  is less than  $m$  and  $|M(\epsilon)| = 0$ .

(3) Let  $\epsilon_1$  and  $\epsilon_2$  be two arbitrary normalized designs given on  $X$ . It is easy to obtain from (1.7.10) that the matrix

$$M(\epsilon) = \alpha M(\epsilon_1) + (1 - \alpha) M(\epsilon_2) \quad (5.1.6)$$

corresponds to the design  $\epsilon = \alpha\epsilon_1 + (1 - \alpha)\epsilon_2$ , that is, belongs to the set under consideration. From this and from the definition of a convex set, the convexity of the set of matrices  $M(\epsilon)$  follows. The closure of this set follows from the closure of  $X$  and the continuity of the functions  $f_{j,\alpha}(x)$  ( $j = 1, 2, \dots, k$ ;  $\alpha = 1, 2, \dots, m$ ).

(4) The validity of Part 4 of the theorem to be proved follows immediately from Caratheodory's theorem which is valid here in view of the fact that the set of all information matrices of a given regression problem  $M(\epsilon)$  is the convex hull of the set of information matrices corresponding to the one-point designs (cf. Section 2.1). The theorem is proved.

The theorem proved permits one to use without any changes many results obtained for the case of a single response surface. For example, Theorems 2.2.4 and 2.2.5 remain valid.

## 5.2. D-Optimal Designs

I. Since the accuracy of the estimator (5.1.4) is characterized only by the covariance matrix defined by (5.1.5), then it is natural to turn to those criteria for comparing experiments which in the one-dimensional case (in the space of resulting observations) were considered in detail in Sections 1.8–1.10.

We consider the properties and methods of constructing  $D$ -optimal designs.

**Lemma 5.2.1.** *For any design  $\epsilon$*

$$\sum_{i=1}^n p_i \operatorname{Tr} \lambda(x_i) d(x_i, \epsilon) = m, \quad (5.2.1)$$

*where  $d(x, \epsilon)$  is the covariance matrix of the estimator  $\hat{\eta}(x, \theta)$  and the summation extends over the entire spectrum of the design  $\epsilon$ .*

*Proof* From (1.7.7) and (1.7.9) it follows that the dispersion matrix of the estimator  $\hat{\eta}(x, \theta)$  is equal to

$$d(x, \epsilon) = M^{-1}(\epsilon) F(x) \quad (5.2.2)$$

By (5.2.2)

$$\sum_{i=1}^n p_i \operatorname{Tr} \lambda(x_i) d(x_i, \epsilon) = \sum_{i=1}^n p_i \operatorname{Tr} \lambda(x_i) F(x_i) M^{-1}(\epsilon) F(x_i)$$

Considering that  $\operatorname{Tr} AB = \operatorname{Tr} BA$ , we obtain

$$\begin{aligned} \sum_{i=1}^n p_i \operatorname{Tr} \lambda(x_i) d(x_i, \epsilon) &= \sum_{i=1}^n \operatorname{Tr} M^{-1}(\epsilon) F(x_i) p_i \lambda(x_i) F(x_i) \\ &= \operatorname{Tr} M^{-1}(\epsilon) M(\epsilon) = \operatorname{Tr} I_m = m, \end{aligned}$$

which is what was required to prove

We now prove the theorem analogous to the theorem of equivalence for the one dimensional case (cf. Theorem 2.2.1)

**Theorem 5.2.1** *The following assertions are equivalent*

- (1) *The design  $\hat{\epsilon}$  maximizes  $|M(\epsilon)|$  (minimizes  $|D(\epsilon)|$ )*,
- (2) *The design  $\hat{\epsilon}$  minimizes  $\max_x \operatorname{Tr} \lambda(x) d(x, \epsilon)$ ,*
- (3)  $\max_x \operatorname{Tr} \lambda(x) d(x, \hat{\epsilon}) = m$

*The information matrices of all designs satisfying conditions 1–3 coincide with one another. Any linear combination of these designs also satisfies 1–3.*

*Proof* (1) We will show that (2) follows from (1). Let the design  $\hat{\epsilon}$  maximize  $|M(\epsilon)|$ . We consider the information matrix corresponding to the linear combination  $\epsilon = (1 - \alpha)\hat{\epsilon} + \alpha\epsilon$ , where  $\epsilon$  is an arbitrary design. By Theorem 5.2.1 such a matrix exists and equals the matrix

$$M(\epsilon) = (1 - \alpha)M(\hat{\epsilon}) + \alpha M(\epsilon)$$

By the definition of the design  $\epsilon$

$$(\partial/\partial\alpha) \log |M(\epsilon)| \Big|_{\alpha=0} = \operatorname{Tr} M^{-1}(\hat{\epsilon}) M(\epsilon) = m$$

Let the design  $\epsilon$  be concentrated at one point  $x$ , then

$$\begin{aligned} \operatorname{Tr} M^{-1}(\hat{\epsilon}) M(\epsilon) &= \operatorname{Tr} M^{-1}(\hat{\epsilon}) F(x) \lambda(x) F(x) \\ &= \operatorname{Tr} \lambda(x) F(x) M^{-1}(\hat{\epsilon}) F(x) = \operatorname{Tr} \lambda(x) d(x, \hat{\epsilon}) \end{aligned}$$

In view of the monotonicity of the function  $\log |M(\epsilon)|$  and the definition of the design  $\hat{\epsilon}$

$$(\partial/\partial\alpha) \log |M(\hat{\epsilon})| |_{\alpha=0} = \text{Tr } \lambda(x) d(x, \hat{\epsilon}) - m \leq 0.$$

On the other hand, from Lemma 5.2.1 it follows that for any design

$$\max_x \text{Tr } \lambda(x) d(x, \epsilon) \geq m.$$

Comparing the last two inequalities, it is easy to obtain the validity of the assertion being proved, in which case

$$\min_{\epsilon} \max_x \text{Tr } \lambda(x) d(x, \epsilon) = \max_x \text{Tr } \lambda(x) d(x, \hat{\epsilon}) = m.$$

(2) We will show that (1) follows from (2).

Suppose the minimax design  $\hat{\epsilon}$  is not *D-optimal*; then, in view of the strict concavity of the function  $\log |M(\epsilon)|$ , a design  $\epsilon$  can be found such that

$$(\partial/\partial\alpha) \log |M(\hat{\epsilon})| |_{\alpha=0} = \text{Tr } M^{-1}(\hat{\epsilon}) M(\epsilon) - m > 0.$$

By Theorem 5.1.1, for any design  $\epsilon$  it is possible to find a design with spectrum consisting of at most  $n = m(m+1)/2 + 1$  points which has the same information matrix. Therefore, without loss of generality, it is possible to assume that the design  $\epsilon$  consists of  $n$  points. In this case,

$$(\partial/\partial\alpha) \log |M(\hat{\epsilon})| |_{\alpha=0} = \sum_{i=1}^n p_i \text{Tr } \lambda(x_i) d(x_i, \hat{\epsilon}) - m > 0. \quad (5.2.3)$$

The summation is taken over the points of the spectrum of the design  $\epsilon$ . But from the minimax property it follows that

$$\text{Tr } \lambda(x_i) d(x_i, \hat{\epsilon}) \leq m, \quad i = 1, 2, \dots, n,$$

or

$$\sum_{i=1}^n p_i \text{Tr } \lambda(x_i) d(x_i, \hat{\epsilon}) - m \leq \sum_{i=1}^n p_i m - m = 0. \quad (5.2.4)$$

The contradiction obtained [cf. (5.2.3) and (5.2.4)] proves our assertion.

(3) The equivalence of assertions (1) and (3), and (2) and (3) immediately follows from (5.2.4) and the equivalence of (1) and (2).

(4) The proof of the concluding part of the theorem is carried out analogously to the proof of the corresponding part of Theorem 2.2.2, and the theorem is proved.

**Corollary 1** At points of the  $D$  optimal design  $\text{Tr } \lambda(x) d(x, \epsilon)$  attains its maximum value  $m$ .

Assume the contrary.

$$\text{Tr } \lambda(x) d(x, \epsilon) < m$$

where  $x$  is one of the points of the spectrum of the design  $\epsilon$ . In view of part (3) of the theorem just proved

$$\sum_{i=1}^n p_i \text{Tr } \lambda(x_i) d(x_i, \epsilon) < \sum_{i=1}^n p_i m - m$$

But according to Lemma 5.2.1

$$\sum_{i=1}^n p_i \text{Tr } \lambda(x_i) d(x_i, \epsilon) = m$$

The obtained contradiction proves our assertion.

The proof of the theorem and its corollary gives a simple method for verifying  $D$  optimality of a design. For this it is sufficient to verify that the quantity  $\text{Tr } \lambda(x) d(x, \epsilon)$  at points of the design equals  $m$  and at the remaining points is at most  $m$ .

**II** The analytical construction of  $D$  optimal designs is possible only in the simplest cases. In this section we present the iterative method of constructing  $D$  optimal designs. This method permits reduction of extremal problems of dimensions  $(k+1)n$  to sequences of extremal problems of the same dimension as the factor space. Here  $n$  is bounded by the limits  $mk$  and  $m(m+1)/2$ . We consider the following iterative procedure:

- 1 There is some nondegenerate design  $\epsilon_0$  with information matrix  $M(\epsilon_0)$ .
- 2 The point  $x_0$  is sought corresponding to

$$\max_x \text{Tr } \lambda(x) d(x, \epsilon_0)$$

## 3. The design

$$\epsilon_1 = (1 - \alpha_0) \epsilon_0 + \alpha_0 \epsilon(x_0)$$

is constructed.

4. The matrix  $M(\epsilon_1)$  is found and its inverse, the covariance matrix  $D(\epsilon_1)$ , is computed.

After this, operations 2–4 are repeated with the design  $\epsilon_1$ , then with  $\epsilon_2$ , and so on as long as one of the following inequalities is not satisfied:

$$\operatorname{Tr} \lambda(x) d(x, \epsilon_s) - m \leq \delta_1, \quad \frac{|M(\epsilon_{s+1})| - |M(\epsilon_s)|}{|M(\epsilon_{s+1})|} \leq \delta_2,$$

where  $\delta_1$  and  $\delta_2$  are some small positive numbers given beforehand.

The step  $\alpha_s$  is chosen according to one of the rules:

- (a)  $\alpha_s$  is the root of the equation

$$(\partial/\partial\alpha) \log |M(\epsilon_{s+1})| = 0 \quad (5.2.5)$$

which is closest to zero;

- (b) the sequence  $\alpha_s$  satisfies the conditions

$$\sum \alpha_s = \infty, \quad \lim_{s \rightarrow \infty} \alpha_s = 0; \quad (5.2.6)$$

- (c)  $\alpha_s$  is divided by  $\gamma > 1$  as soon as

$$|M(\epsilon_s)| > |M(\epsilon_{s+1})|. \quad (5.2.7)$$

In the computation of the matrix  $D(\epsilon_{s+1})$  and the search for the solution, for the case when the number  $k$  of simultaneously observed quantities is significantly less than the number of unknown parameters  $m$ , it is useful to use the following formulas:

$$D(\epsilon_{s+1}) = (1 - \alpha_s)^{-1} \left\{ I_m - \frac{\alpha_s}{1 - \alpha_s} D(\epsilon_s) F(x_s) \times \left[ I_k + \frac{\alpha_s}{1 - \alpha_s} \lambda(x_s) F'(x_s) D(\epsilon_s) F(x_s) \right]^{-1} \lambda(x_s) F(x_s) \right\} D(\epsilon_s), \quad (5.2.8)$$

$$|D(\epsilon_{s+1})| = (1 - \alpha_s)^{-m} \left| I_k + \frac{\alpha_s}{1 - \alpha_s} \lambda(x_s) F'(x_s) D(\epsilon_s) F(x_s) \right|^{-1} |D(\epsilon_s)|. \quad (5.2.9)$$

Formula (5.2.8) reduces the operation of inversion of the matrix of dimension  $m \times m$  to the operation of inversion of a matrix of dimension  $k \times k$ , formula (5.2.9) reduces the computation of the determinant of order  $m$  to the computation of the determinant of order  $k$ .

**Theorem 5.2.2.** *The iterative process 1-4 converges, in which case*

$$\lim_{s \rightarrow \infty} |M(\epsilon_s)| = |M(\hat{\epsilon})|,$$

where  $\hat{\epsilon}$  is a D-optimal design

*Proof* The proof is carried out for  $\alpha_s$  chosen by (5.2.6). Generalization of the proof to the remaining cases can be carried out by the reader without difficulty.

By Theorem 5.2.1

$$\operatorname{Tr} \lambda(x_s) d(x_s, \epsilon_s) > m$$

if the design  $\epsilon_s$  is distinct from the D-optimal. From the definition of  $\epsilon_{s+1}$ ,

$$(\partial/\partial\alpha) \ln |M(\epsilon_{s+1})| \Big|_{\alpha=0} = \operatorname{Tr} \lambda(x_s) d(x_s, \epsilon_s) - m > 0$$

If  $\alpha_s$  is chosen as shown in the beginning of the proof of the theorem, then from the preceding inequalities it follows that

$$|M(\epsilon_s)| \leq |M(\epsilon_{s+1})|$$

The monotone nondecreasing sequence  $\{|M(\epsilon_s)|\}$  is bounded since  $|M(\hat{\epsilon})| \geq |M(\epsilon)|$  for any design  $\epsilon$ . Therefore it converges to some limit  $|M(\epsilon)|$ .

We show that  $|M(\epsilon)| = |M(\hat{\epsilon})|$ . From this, in view of Theorem 5.2.1, it will also follow that  $M(\epsilon) = M(\hat{\epsilon})$ .

Assume the contrary  $|M(\epsilon)| \neq |M(\hat{\epsilon})|$ . Then (cf. Theorem 5.2.1)

$$\lim_{s \rightarrow \infty} [\operatorname{Tr} \lambda(x_s) d(x_s, \epsilon_s) - m] = \Delta > 0,$$

or, by the definition of  $\alpha_s$ ,

$$\lim_{s \rightarrow \infty} [|M(\epsilon_{s+1})| - |M(\epsilon_s)|] = \delta(\Delta) > 0$$

This contradicts the condition of convergence of the sequence  $\{|M(\epsilon_s)|\}$ . In this manner

$$\lim_{s \rightarrow \infty} |M(\epsilon_s)| = |M(\hat{\epsilon})|.$$

The theorem is proved.

EXAMPLE 1. At each point  $x$  let observations of two uncorrelated quantities  $y_1$  and  $y_2$  be possible:

$$\begin{aligned}\eta_1(x, \theta) &= \theta_1 + \theta_2 x + \theta_3 x^2, & 0 \leq x \leq 1 \\ \eta_2(x, \theta) &= \theta_4 x + \theta_5 x^3 + \theta_6 x^4.\end{aligned}\quad (5.2.10)$$

Also, let the errors for observations of the two quantities be equal and not depend on  $x$ :

$$\lambda(x) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

As an initial approximation the following design is chosen:

$$\epsilon_0 = \left\{ \begin{array}{l} x_1 = 0.0; x_2 = 0.20; x_3 = 0.40; x_4 = 0.60; x_5 = 0.80; x_6 = 1.0 \\ p_i = \frac{1}{6}, \quad i = 1, 2, \dots, 6 \end{array} \right\}.$$

After 20 iterations the design  $\epsilon_{20}$  was obtained:

$$\epsilon_{20} = \left\{ \begin{array}{l} x_1 = 0.0; \quad x_2 = 0.38; \quad x_3 = 0.76; \quad x_4 = 1.0 \\ p_1 = 0.16; \quad p_2 = 0.28; \quad p_3 = 0.23; \quad p_4 = 0.33 \end{array} \right\}.$$

In the design  $\epsilon_{20}$  points with  $p_i < 0.01$  were dropped. The maximal sum of the dispersions is equal to

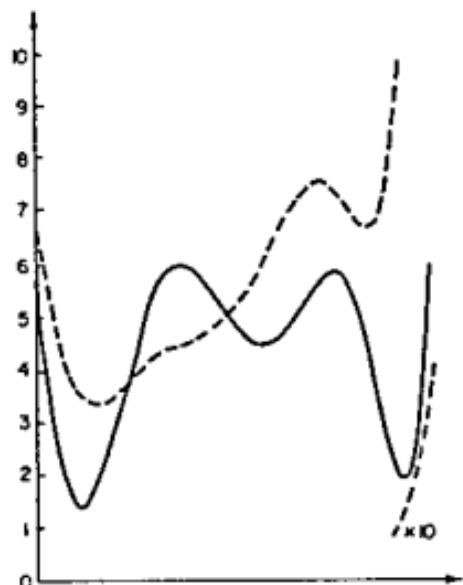
$$\max_x \text{Tr } \lambda(x) d(x, \epsilon_{20}) = 6.01.$$

We note that for the exact  $D$ -optimal design

$$\max_x \text{Tr } \lambda(x) d(x, \hat{\epsilon}) = m = 6.$$

It is interesting to compare the characteristics of the design  $\epsilon_{20}$  and the conventional equiweighted designs, for example, the design  $\epsilon_0$ . For the indicated equiweighted design (cf. Fig. 22):

$$\max_x \text{Tr } \lambda(x) d(x, t_0) = 39.$$



**Fig. 22.** Regression problem (5 2 10)  
The sum of the variances  $d_1(x)$  for the  
design  $e_{10}$  is plotted by the solid line. The dashed line plots the same  
quantity for the uniform design

In (5 2 10) the quantities  $y_1$  and  $y_2$  depend on different parameters. We consider the case when  $y_1$  and  $y_2$  depend on the same parameters:

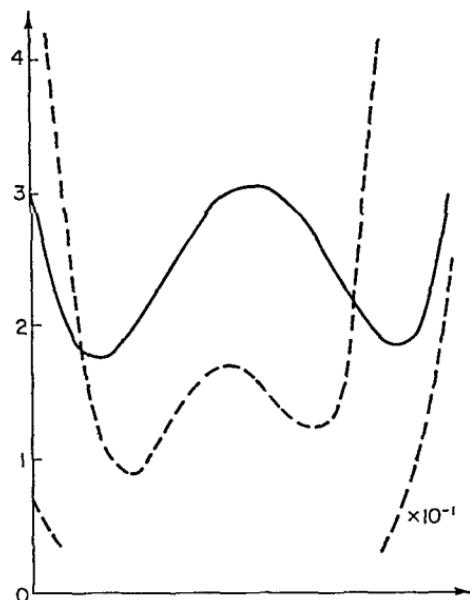
$$\begin{aligned}\eta_1(x, \theta) &= \theta_1 + \theta_2 x + \theta_3 x^2, & 0 \leq x \leq 1 \\ \eta_2(x, \theta) &= \theta_1 + \theta_2 x^2 + \theta_3 x^4\end{aligned}\quad (5 2 11)$$

The corresponding  $D$ -optimal design, determined by the iterative procedure presented, is equal to (cf Fig. 23)

$$\epsilon = \left\{ \begin{array}{l} x_1 = 0.0, \quad x_2 = 0.54, \quad x_3 = 1.0 \\ p_1 = \frac{1}{3}, \quad p_2 = \frac{1}{3}, \quad p_3 = \frac{1}{3} \end{array} \right\}$$

**III.** The results of Parts I and II can be generalized to the case of nonlinear parametrization. In this case it is necessary to consider locally  $D$ -optimal designs (cf Section 2 8). Replacing formulas where necessary by those introduced in these sections, the reader can independently carry out these generalizations without difficulty by relying, for example, on the corresponding exposition of the example considered below.

**EXAMPLE 2** We consider a chemical reaction of the type  $A \rightarrow B \rightarrow C$ . We assume that all reactions are of the first order [48]. This indicates



**Fig. 23.** Regression problem (5.2.11).  
The sum of the variances  $d_i(x)$  for the approximate  $D$ -optimal design is plotted by the solid line. The dashed line plots the same quantity for the uniform design.

that the rate of reaction of the substance  $A$  is directly proportional to its concentration with a constant of proportionality equal to  $\theta_1$ ; the rate of formation of  $C$  is directly proportional to the concentration of  $B$  with the constant of proportionality  $\theta_2$ ; the rate of formation of  $B$  is proportional to the concentration of  $A$  with constant of proportionality  $\theta_1$ , and the rate of decrease of  $B$  is proportional to the concentration of  $B$  with constant of proportionality  $\theta_2$ . It is possible to show that the reaction is described in the following terms by means of the response surfaces

$$\eta_1(x, \theta) = e^{-\theta_1 x},$$

$$\eta_2(x, \theta) = (e^{-\theta_1 x} - e^{-\theta_2 x}) \theta_1 / (\theta_2 - \theta_1),$$

$$\eta_3(x, \theta) = 1 + (-\theta_2 e^{-\theta_1 x} + \theta_1 e^{-\theta_2 x}) / (\theta_2 - \theta_1),$$

where  $\eta_1$ ,  $\eta_2$ ,  $\eta_3$  indicate the average concentration of the substances  $A$ ,  $B$ , and  $C$  respectively.

We will assume that simultaneous observations of the quantities  $y_1$  and  $y_2$  ( $k = 2$ ) are possible and that the observations taken from different points of the factor space are independent.

Suppose that it is known by some considerations that the true values of the parameters  $\theta_1$ ,  $\theta_2$  are found close to  $\theta_{1(0)} = 0.7$  and

$\theta_{2(0)} = 0.2$  and that the covariance matrix of the observations  $y_1(x)$  and  $y_2(x)$  is constant, in the region where the observations are taken, and equal to

$$d = \begin{vmatrix} 1 & 1 \\ 1 & 4 \end{vmatrix} \quad \text{or} \quad \lambda(x) = \begin{vmatrix} 1.333 & -0.333 \\ -0.333 & 0.333 \end{vmatrix}$$

We determine the functions

$$f_{ka} = \partial \eta_k(x, \theta) / \partial \theta_a, \quad k = 1, 2, \quad a = 1, 2$$

Carrying out the differentiation, we obtain

$$f_{11}(x) = -xe^{-\theta_1 x}, \quad f_{12}(x) = 0,$$

$$f_{21}(x) = (\theta_2 - \theta_1)^{-2} \{ \theta_2(e^{-\theta_2 x} - e^{-\theta_1 x}) + \theta_1(\theta_2 - \theta_1)x e^{-\theta_1 x} \}$$

$$\alpha, \beta = 1, 2, \quad \alpha \neq \beta$$

The elements of the information matrix have the form

$$M_{ab} = \sum_{i=1}^n p_i \sum_{k=1}^2 \lambda_{ki} f_{ka}(x_i) f_{ib}(x_i)$$

The design obtained by the iterative procedure of Part III of this section has the following characteristics

$$\hat{\epsilon}(\theta_0) = \begin{cases} x_1 = 1.75 & x_2 = 4.80 \\ p_1 = 0.8 & p_2 = 0.2 \end{cases}$$

$$D(\theta_0) = 1.29$$

The design obtained by direct minimization of  $|D(\theta_0)|$  under the assumption  $p_1 = p_2 = 0.5$  (cf [48]) has a significantly larger value of the determinant

$$|D(\theta_0, p_1 = p_2)| = 1.67$$

In this case

$$\hat{\epsilon}(\theta_0, p_1 = p_2) = \begin{cases} x_1 = 1.4 & x_2 = 6.8 \\ p_1 = 0.5 & p_2 = 0.5 \end{cases}$$

**IV.** Generalizations of the result obtained for one-dimensional  $y$  are possible also for the truncated  $D$ -criteria. The methods of generalization are the same as for the  $D$ -criteria. We differentiate  $(\partial/\partial\alpha) \log |D_{ll}||_{\alpha=0}$ , which plays a basic role in the formulations of the theorems analogous to Theorem 2.7.1. It is easily computed by formula (2.7.4) and equals

$$(\partial/\partial\alpha) \log |D_{ll}||_{\alpha=0} = \text{Tr } \lambda(x) d(x, l, \epsilon) - l, \quad (5.2.12)$$

where  $l$  is the number of parameters presenting interest to the experimenter, and

$$d(x, l, \epsilon) = F'(x) \begin{vmatrix} D_{ll} & D_{lk} \\ D_{kl} & D_{kk} - M_{kk}^{-1} \end{vmatrix} F(x).$$

Relying on (5.2.12) and repeating the arguments analogous to those carried out in Parts I–III, it is not difficult to obtain that all results of Section 2.7 remain valid if the quantity  $\lambda(x) d(x, l, \epsilon)$  is replaced by  $\text{Tr } \lambda(x) d(x, l, \epsilon)$ .

### 5.3. Linear-Optimal Designs

**I.** When observations on several response surfaces are simultaneously possible, the methods of constructing linear-optimal designs (obtained for one-dimensional  $y$ ) coincide in the large with methods given in the preceding section. Therefore some of the proofs given in what follows and strings of computations either are omitted or are carried out in abbreviated form.

The proof of the basic theorem on properties of linear-optimal designs for simultaneous observations of several response surfaces depends on the following assertion.

**Lemma 5.3.1.** *For any design  $\epsilon$*

$$\sum_{i=1}^n p_i \varphi(x_i, \epsilon) = L[D(\epsilon)], \quad (5.3.1)$$

where  $D(\epsilon)$  is the covariance matrix of the estimator  $\hat{\theta}$  for the design  $\epsilon$  and

$$\varphi(x, \epsilon) = L[D(\epsilon) F(x) \lambda(x) F(x) D(\epsilon)].$$

*Proof* From (2.9.2) and (2.9.3) and the definition of the information matrix it follows that

$$\begin{aligned} \sum_{i=1}^n p_i L[D(\epsilon) F(x_i) \lambda(x_i) F'(x_i) D(\epsilon)] &= L\left[D(\epsilon) \left(\sum_{i=1}^n p_i F(x_i) \lambda(x_i) F'(x_i)\right) D(\epsilon)\right] \\ &= I[D(\epsilon) M(\epsilon) D(\epsilon)] \\ &= L[D(\epsilon)] \end{aligned}$$

The lemma is proved

**Lemma 5.3.2** Let  $\epsilon = (1 - \alpha)\epsilon + \alpha\epsilon(x)$ , where the design  $\epsilon(x)$  is concentrated at the single point  $x$ . Then

$$(\partial/\partial\alpha) L[D(\epsilon)]|_{\alpha=0} = L[D(\epsilon)] - \varphi(x, \epsilon) \quad (5.3.2)$$

*Proof* From Lemma 2.9.2 and the definition of the functional  $L$ ,

$$\begin{aligned} (\partial/\partial\alpha) L[D(\epsilon)]|_{\alpha=0} &= -L\{D(\epsilon)[M(\epsilon(x)) - M(\epsilon)] D(\epsilon)\} \\ &\quad - L[D(\epsilon)] + L\{D(\epsilon) M[\epsilon(x)] D(\epsilon)\} \\ &= L[D(\epsilon)] - \varphi(x, \epsilon) \end{aligned}$$

which proves the lemma

Let  $L$  satisfy (2.9.2)–(2.9.4). Then relying on the lemma just proved and repeating the proof of Theorem 2.9.2 practically word for word (cf. also Section 5.2) it is not difficult to obtain the validity of the following theorem

**Theorem 5.3.1** The following assertions are equivalent

- (1) the design  $\hat{\epsilon}$  minimizes  $L[D(\epsilon)]$
- (2) the design  $\hat{\epsilon}$  minimizes  $\max_x \varphi(x, \epsilon)$
- (3)  $\max_x \varphi(x, \epsilon) = L[D(\epsilon)]$

Any linear combination of designs satisfying conditions 1–3 also satisfies these conditions

If  $L(A) > 0$  where  $A$  is a positive semidefinite matrix, then the strong variant of Theorem 5.3.1 (cf. Section 2.9) holds

**Theorem 5.3.1a** The following assertions are equivalent

- (1) the design  $\hat{\epsilon}$  minimizes  $L[D(\epsilon)]$ ,

- (2) the design  $\tilde{\epsilon}$  minimizes  $\max_x \varphi(x, \tilde{\epsilon})$ ,  
 (3)  $\max_x \varphi(x, \tilde{\epsilon}) = L[D(\tilde{\epsilon})]$ .

The information matrices of all designs satisfying 1–3 coincide with one another. Any linear combination of designs satisfying conditions 1–3 also satisfies these conditions.

**II.** As a numerical method of constructing linear-optimal designs it is possible to use an iterative procedure analogous to procedure 1–4 of the preceding section.

1. Some nondegenerate design  $\epsilon_0$  is available with dispersion matrix of the unknown parameters  $D(\epsilon_0)$ .
2. The point  $x_0$  is sought where  $\max_x \varphi(x, \epsilon_0)$  is attained.
3. The design  $\epsilon_1 = (1 - \alpha_0)\epsilon_0 + \alpha_0\epsilon(x_0)$  is constructed.
4. The matrix  $D(\epsilon_1)$  is found.

After this, operations 2–4 are repeated with the design  $\epsilon_1$ , afterward with  $\epsilon_2$ , and so on as long as the following inequality is not satisfied:

$$\{L[\epsilon_s] - L[\epsilon_{s+1}]\}/L[\epsilon_{s+1}] \leq \delta.$$

The step  $\alpha_s$  is chosen either from consideration of the maximal decrease of  $L[\epsilon_s]$  for a given  $x_s$ , or the sequence  $\{\alpha_s\}$  is chosen as specified in (5.2.6) and (5.2.7).

It is possible to show that (cf. Section 5.2) the iterative process converges:

$$\lim_{s \rightarrow \infty} L(\epsilon_s) = L(\tilde{\epsilon}).$$

If the design  $\tilde{\epsilon}$  is nondegenerate, then

$$L(\tilde{\epsilon}) = \min_{\epsilon} L(\epsilon).$$

In the remaining cases

$$L(\epsilon_0) > L(\tilde{\epsilon}) \geq \min_{\epsilon} L(\epsilon).$$

The values of  $\varphi(x, \epsilon)$  for various criteria are presented in Table 8. In those cases when  $k \ll m$ , where  $k$  is the number of response surfaces and  $m$  is the number of unknown parameters, it is useful to use the representation (5.2.8) for computation of the matrix  $D(\epsilon_{s+1})$ .

Table 8

Values of  $\varphi(x, \epsilon)$  for Various Criteria

$L[D(\epsilon)]$	$\varphi(x, \epsilon)$
$\text{Tr } D(\epsilon)$	$\text{Tr } \lambda(x) F(x) D(\epsilon) F(x)$
$\text{Tr } d(x_0 - \epsilon)$	$\text{Tr } d(x - x_0 - \epsilon) \lambda(x) d(x - x_0 - \epsilon)$ where $d(x - x_0 - \epsilon) = F(x) D(\epsilon) F(x_0)$
$\int_{\mathcal{Z}} \text{Tr } d(x - \epsilon) dx$	$\text{Tr } \lambda(x) F(x) D(\epsilon) \bar{M} D(\epsilon) F(x)$ where $\bar{M} = \int_{\mathcal{Z}} F(x) F(x) dx$
$D(\theta) = D(I \theta)$ where $I = \ I_1 \ I_2 \ \dots \ I_m\ $	$q(x) \lambda(x) q(x)$ where $q(x) = I D(\epsilon) F(x)$
$E[(\theta - \hat{\theta}) A(\theta - \hat{\theta})] = \text{Tr } A D(\epsilon)$ where $A = CC'$	$\text{Tr } \lambda(x) F(x) D(\epsilon) A D(\epsilon) F(x)$

#### 5.4 Sequential Design

The results of the two preceding sections permit generalization of the methods of sequential designs developed in Chapter 4 to the case of simultaneous measurements of several response surfaces. In all calculations it is sufficient to replace

$$\begin{aligned} \lambda(x) d(x | N) &\quad \text{by} \quad \text{Tr } \lambda(x) d(x | N) \\ \lambda(x) L[D(N) f(x) f'(x) D(N)] &\quad \text{by} \quad L[D(N) F(x) \lambda(x) F'(x) D(N)] \end{aligned}$$

The generalizations of the lemmas and theorems analogous to those presented in Sections 5.2 and 5.3 can be set up independently by the reader.

In carrying out the computations, necessary for sequential design the following formulas appear to be very useful (for  $k < m$ )

$$\begin{aligned} D(N + \Delta N | x) &= \{I_m - \Delta N D(N) F(x) \\ &\quad \times [I_k + \Delta N \lambda(x) F(x) D(N) F(x)]^{-1} \lambda(x) F(x)\} D(N) \end{aligned} \quad (5.4.1)$$

and

$$|D(N + \Delta N, x)| = |D(N)| / |I_k + \Delta N \lambda(x) F'(x) D(N) F(x)|. \quad (5.4.2)$$

The proof of formula (5.4.1) depends on Lemma 2.6.1, and the proof of formula (5.4.2) on Lemma 2.5.1. The arguments proceed in the same way as for the case of a single response surface.

# 6

## Discriminating Experiments

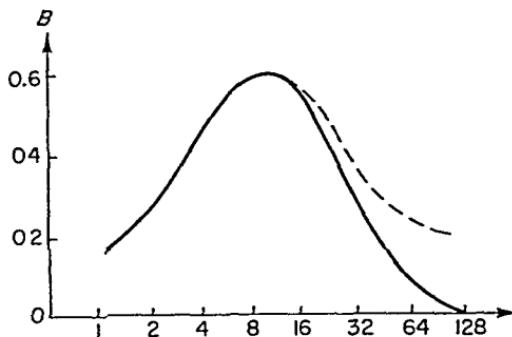
### 6.1. Statement of the Problem

I In the definitive stage of a project the experimenter is frequently faced with the situation that existing data, experimental or theoretical, satisfy two or more mathematical models Before proceeding to further investigations (for example, the precise estimation of several parameters) it is necessary to set up an experiment that would permit discrimination between the models The design of a suitable experiment consists in finding those points at which the comparison of models would be set up under optimal conditions In other words, it is necessary to find points for which the resulting observations are far from being invariant with respect to a change from one mathematical model to another We clarify this with an example

Consider a chemical reaction in which a substance  $A$  is used for obtaining a product  $B$  We assume that the experimenter knows that the reaction can be one of two forms



Typical curves indicating the dependence of the concentration of the product as a function of time are given in Fig. 24 It is clear that in



**Fig. 24.** Concentration of the product  $B$  as a function of time for the reaction  $A \rightarrow B \rightarrow C$  (solid line) and the reaction  $A \rightarrow B \rightleftharpoons C$  (dashed line).

order to clarify which of the models is true, it is completely pointless to take observations in the region  $t \leq 10$  min.

Increased information on which of the models is true can be obtained by measurements taken at larger times. However, measurements at times which are too large are economically disadvantageous; therefore it is necessary to find some compromise solution. In this given case, the discriminating observations evidently should be taken in the region  $t \sim 100$  min.

We consider the example as being comparatively simple. In practice, the search for the most useful region for distinguishing concurring models is, as a rule, a complicated problem and requires the application of special mathematical techniques, the exposition of which is relegated to the subsequent sections of this chapter. In some cases, one succeeds in combining a discriminating experiment with an experiment in determining the unknown parameters; in others, it is necessary to set up a special discriminating experiment. There is no sharp boundary between these cases, and only the experimenter himself, considering the real situation, can decide which type of experiment to carry out.

**II.** From the point of view of mathematical statistics, the problem considered in this chapter is formulated in the following manner:

Some quantity  $y$  is measured. The results of the measurement are random quantities, which satisfy the requirements enumerated below.

- At each point  $x$  of the factor space  $X$  the random quantity  $y$  is distributed with conditional density function

$$p_j(y | x) = p_j[y | \eta_j(x, \theta_j)], \quad j = 1, 2, \dots, v.$$

The analytic form of the conditional density function and the function  $\eta_j(x, \theta_j)$  is assumed known. The sets of parameters  $\theta_j$  in the general case can be unknown.

- 2 The mean value of the results of observations is equal to

$$E(y|x) = \eta_j(x, \theta_j)$$

- 3 For the majority of the methods outlined in what follows for designing discriminating experiments, it will be necessary to know the efficiency function of the experiment

$$\lambda_j(x) = D_j^{-1}(y|x) = \left\{ \int_Y [y - \eta_j(x, \theta_j)]^2 p_j(y|x) dy \right\}^{-1},$$

where  $Y$  is the domain of values for  $y$ .

- 4 If not specifically stated, then the results of the observations are assumed to be independent.

The set of requirements 1–4 for a given  $j$  will be called hypothesis  $H_j$ .

Let the set of concurring hypotheses  $H_1, H_2, \dots, H_v$  contain the true hypothesis (that is, the one corresponding to the real process being carried out). Such a collection of hypotheses will be called complete. If the experimenter has no guarantee that the true hypothesis belongs to the collection  $H_1, H_2, \dots, H_v$ , then the specified set will be called incomplete.

We consider several concepts which are fundamental in constructing criteria for discriminating hypotheses formulated in the form of 1–4.

Let the parameters  $\theta_j$  ( $j = 1, 2, \dots, v$ ) be given *a priori* ( $\theta_j = \theta_{j0}$ ,  $j = 1, 2, \dots, v$ ). In this case the hypotheses  $H_j$  are called simple.

If the parameters  $\theta_j$  are not given and it is only known that they belong to some domain  $\Omega_j$  ( $\theta_j \in \Omega_j$ ,  $j = 1, 2, \dots, v$ ), then the hypotheses are called composite.

We assume that the collection of hypotheses  $H_1, H_2, \dots, H_v$  is complete and that for the experiment  $\mathcal{E}$  the observations  $y_1, y_2, \dots, y_n$  were obtained at the points  $x_1, x_2, \dots, x_n$  (some points  $x_i$  can coincide with one another).

The process of seeking the true hypothesis must depend on some decision rule  $\delta$ , corresponding to which the collection of all possible samples  $Y_n$  (cf. Section 1.2) is divided into  $v$  nonintersecting regions  $R_{nj}$  ( $j = 1, 2, \dots, v$ ). If the sample  $Y_n$  belongs to the region  $R_{nj}$ , then hypothesis  $H_j$  is chosen. Sometimes, along with the sets

$R_{nj}$  ( $j = 1, 2, \dots, v$ ) there is also introduced a so-called indifference zone  $R_{n0}$ . If the sample  $Y_n$  belongs to  $R_{n0}$ , none of the hypotheses  $H_j$  is chosen and the observations are continued. The region  $\bar{R}_{nj}$ , composed of the regions  $R_{n1}, \dots, R_{n(j-1)}, R_{n(j+1)}, \dots, R_{nv}$ , is called the critical region with respect to the hypothesis  $H_j$  ( $\bar{R}_{nj} = \bigcup_{k \neq j} R_{nk}$ ).

The results of each experiment can have a loss  $Q(\delta)$ , which is incurred from the allocation for carrying out the observations and the possible penalty in case a wrong decision is made. If we realize that the quantities  $y_1, y_2, \dots, y_n$  are random, then generally speaking, we will obtain various values of the loss from experiment to experiment for a given decision rule  $\delta$  and a given sample size. It is natural to require that the decomposition of the region into  $R_{n0}, R_{n1}, \dots, R_{nv}$  is made in such a way that the expected value with respect to  $Y_n$  of the loss is minimized:

$$\mathcal{R}(\delta) = E_{Y_n}[Q(\delta)]. \quad (6.1.1)$$

Since the distribution function  $\Psi(Y_n)$  depends on the form of the true hypothesis, then the expected loss is a function of  $H_T$ . The function  $R(\delta)$  is called the risk function.

In the general case, one cannot find a rule  $\tilde{\delta}$  that would minimize  $\mathcal{R}(\delta)$  without dependence on the value  $H_T$  ( $\tilde{\delta}$  is called a rule with uniformly minimum risk [49]). Therefore it is necessary either to find a rule with uniformly minimum risk for some reduced class of hypothesis or to find a rule minimizing the Bayes risk.

By the Bayes risk we understand the quantity

$$r[p_0(H), \delta] = E_H[\mathcal{R}(\delta)], \quad (6.1.2)$$

where the operator  $E_H$  indicates the expectation with respect to the collection of hypotheses  $H$  with *a priori* density function  $p_0(H)$ . In the absence of prior information concerning the true hypothesis it is possible to minimize the maximum risk

$$\max_H \mathcal{R}(\delta). \quad (6.1.3)$$

In the future, without considering the criterion of optimality used for choosing the optimal rule, we will indicate the optimal rule by  $\delta$ , and the function which  $\delta$  minimizes by  $r(\delta)$ .

It is not difficult to see that for the hypotheses  $H_j$  ( $j = 1, 2, \dots, v$ ), formulated in the form 1–4, the function  $r(\delta)$  depends not only on the rule  $\delta$  for constructing the regions  $R_{n0}, R_{n1}, \dots, R_{nv}$ , but also on the

location of the points  $x_1, x_2, \dots, x_n$  in the factor space [on the design  $\mathcal{E}(n)$ ] Therefore it makes sense to talk about the design of a discriminating experiment, i.e., about the search for designs  $\mathcal{E}(n)$  which minimize  $r[\delta, \mathcal{E}(n)]$

**III** If the collection of hypotheses  $H_j$  ( $j = 1, 2, \dots, v$ ) is incomplete, then statements of the type, "we accept the hypothesis  $H_j$ ," are impossible. The experimenter can make a statement only that the results of the observation do not contradict  $H_{j_1}, H_{j_2}, \dots, H_{j_k}$  ( $k \leq v$ ), or the opposite statement that the results of the observations contradict hypotheses  $H_{j_1}, H_{j_2}, \dots, H_{j_k}$  and these hypotheses may be dropped. In the case of an incomplete collection of the concurring hypotheses, the assertion that the hypotheses  $H_1, \dots, H_{j-1}, H_{j+1}, \dots, H_v$  are not true, as distinct from the case when the collection of concurring hypotheses is complete, is not equivalent to the assertion that the hypothesis  $H_j$  is true. Evidently, for the case considered it makes sense to talk only about the critical regions  $R_{nj}$  ( $j = 1, 2, \dots, v$ ) which, generally speaking, can intersect one another, and the region of indifference  $R_{n0}$ .

The stated remarks are taken into account in constructing the loss function  $Q(\delta)$ , where the rule  $\delta$  specifies the method of constructing the regions  $R_{nj}$ .

We note also that the density  $p_0(H)$  can be defined only for a complete collection of hypotheses. Therefore, in the case under consideration one cannot rely on the Bayes risk [cf. (6.1.2)].

**IV.** Summarizing the material from Parts II and III, we enumerate several basic problems which the experimenter comes in contact with in discriminating hypotheses.

1 The choice of the collection of concurring hypotheses. The strongest results can be obtained if one succeeds in finding a complete system of concurring hypotheses.

It is evident that the design will be more effective the smaller the size of the complete collection of concurring hypotheses.

2 Construction of the loss function  $Q(\delta)$ . The function  $Q(\delta)$  is usually sought as some compromise between two concurring requirements which are difficult to satisfy in the majority of cases. The function must have a simple analytic form and at the same time must describe the real loss, accompanying the experiment  $\mathcal{E}$  and the decision rule  $\delta(\mathcal{E})$ , sufficiently well.

3. In using the Bayes approach it is necessary to analyze carefully the prior information for constructing the *a priori* probabilities  $p_0(H_j)$ .
4. The choice of an optimum decision rule for a given function  $r(\delta)$ .
5. The optimal allocation of resources in the region of action  $X$  (the design of the experiment).

The solution of the problems outlined in 1 and 3 rest basically on the shoulders of the experimenter as a qualified specialist in that region of science which corresponds to the investigation being carried out. Only a specialist in this situation can extract the most information from theoretical and analogous experimental investigations and afterwards present the information in terms of a small number of hypotheses  $H_j$  and their *a priori* probabilities  $p_0(H_j)$ .

The choice of a suitable loss function usually centers on those classes of functions  $Q(\delta)$  for which the problem of choosing an optimal decision rule  $\delta$  has a solution.

The most complicated, from the mathematical point of view, are problems presented in 4 and 5.

The problem of choosing an optimal decision rule is treated in a broad literature (cf., for example, [49, 50]). Significantly small attention is paid in the literature to designing discriminating experiments.

The current chapter basically is connected with the construction of the mathematical techniques of designing discriminating experiments. In the majority of cases the decision rule will be assumed given. Information about the properties of the decision rule being used can be obtained in the monographs indicated above.

Sometimes, in order to emphasize that the discussion is about discriminating hypotheses of a special form (cf. Part II), instead of the term "discriminating hypotheses" the term "discriminating mathematical models of a process" will be used or "seeking the best mathematical model of the process."

## 6.2. Criteria Depending on the Difference of Sums of Weighted Squares of Residuals

1. We assume now that the results of the observations are independent, normally distributed random variables

$$p_{\eta}[\gamma | \eta_j(x, \theta_j)] = (2\pi)^{-1/2} b^{-1}(x) \exp\left\{-\frac{1}{2}\left[\frac{\gamma - \eta_j(x, \theta_j)}{b(x)}\right]^2\right\}. \quad (6.2.1)$$

Here and in what follows we will deal only with normal distributions, although in practice the results of observations are only in rare cases obtained exactly from the normal distribution. There is a series of arguments on the usefulness of separating the normal law from the remaining distribution laws.

1 The basic role of the normal distribution is indicated by the central limit theorem [11, 12]. Applying this theorem, in particular, to the results of observations grouped at one point  $x_i$  of the factor space, it is possible to obtain that their arithmetic mean  $y_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$  has a distribution close to normal for sufficiently large  $n_i$ .

Grouping the observations and afterwards using  $y_i$  as the result, we can apply any method developed under the assumption of the normal law of distribution for the results of observations. Because of this, part of the information "is lost" in some cases, that is, the statistic depending on  $y_i$  and  $n_i^{-1} b_i^2$  can be insufficient [11]. In the majority of practical cases these losses can be neglected.

Since the construction of the best linear estimates requires only the variables  $y_i$  and their dispersions  $n_i^{-1} b_i^2$  (cf. Corollary 6 of Theorem 1.3.2), then it is not difficult to obtain that the best linear estimates have in this case a normal distribution. Relying on the central limit theorem, it is also possible to verify that the best linear estimate has a distribution law close to normal also for ungrouped observations, if in general the number of observations is sufficiently large ( $N \rightarrow \infty$ ).

2 The analytic form of the function (6.2.1) is simple and these functions are well known. Detailed studies and tabulation of the distribution of probabilities for a broad class of quantities are available for normally distributed random variables.

3 The normal distribution well describes asymptotically many distributions, for example, the Poisson  $\chi^2$  distribution (cf. [12, 51]).

II. Suppose there are two concurring hypotheses  $H_1$  and  $H_2$ , formulated in accordance with Part II of the preceding section. We consider the following decision rule:

After  $N$  observations

I hypothesis  $H_1$  is accepted if

$$S_2(N) - S_1(N) > 0 \quad (6.2.2)$$

where

$$S_j(N) = \sum_{i=1}^n w_i [y_i - \eta_j(x_i, \hat{\theta}_j)]^2, \quad j = 1, 2, \quad (6.2.3)$$

$\hat{\theta}_j$  is the best linear (quasi linear) estimate under the assumption of the truth of the corresponding hypothesis, and  $N$  is the total number of observations taken at the points  $x_1, x_2, \dots, x_n$ ;

2. hypothesis  $H_2$  is accepted if

$$S_2(N) - S_1(N) < 0. \quad (6.2.4)$$

We assume that the loss for falsely accepting the  $k$ th hypothesis equals  $\gamma_k$ . If the  $k$ th hypothesis is accepted, when the  $j$ th hypothesis is true, then the loss will be equal to

$$Q_j = cN + \gamma_k, \quad (6.2.5)$$

where  $c$  is the cost of each of the  $N$  observations.

Let the number of observations  $N$  be fixed and given *a priori*. Then the Bayes risk for a given decision rule and loss (6.2.5) equals

$$r(N) = cN + \sum_{\substack{j,k=1 \\ j \neq k}}^2 p_0(H_j) \gamma_k P[S_j(N) - S_k(N) > 0 | H_j]. \quad (6.2.6)$$

If the analytic form of the density function  $p_j[x | \eta_j(x, \theta_j)]$  is known, then the conditional probability

$$P[S_j(N) - S_k(N) > 0 | H_j] \quad (6.2.7)$$

can in principle be computed for each of the designs  $\mathcal{E}(N)$ . However, even for comparatively simple response surfaces, it is necessary to deal with serious computational difficulties which make construction of the optimal design practically impossible.

This situation can be remedied by changing the function  $r(N)$  to some other function  $\tilde{r}(N)$  close to  $r(N)$  in the region of interest to the experimenter and at the same time not requiring cumbersome computations. More simple and useful results are obtained in those cases when the probability (6.2.7) can be approximated by some function

$$P_j\{E_j[S_j(N) - S_k(N)]\} \quad (j, k = 1, 2, \quad j \neq k),$$

where  $E_j$  is the expectation operator with respect to results of observa-

tions under the assumption that the  $j$ th hypothesis is true. In this situation the Bayes risk will be closely approximated by the function

$$\tilde{r}(N) = cN + \sum_{\substack{j, k=1 \\ j \neq k}}^n p_0(H_j) \gamma_k P_j(E_j[S_j(N) - S_k(N)]) \quad (6.2.8)$$

The approximation (6.2.8) is usually sufficient for large  $E_j[S_j(N) - S_k(N)]$ .

The design of the discriminating experiment under the assumption of the closeness of  $\tilde{r}(N)$  and  $r(N)$  consists of seeking a design  $\delta(x_1, x_2, \dots, x_n, p_1, p_2, \dots, p_n)$  minimizing the quantity

$$l(N) = \sum_{\substack{j, k=1 \\ j \neq k}}^n p_0(H_j) \gamma_k P_j(E_j[S_j(N) - S_k(N)]) \quad (6.2.9)$$

The necessity of seeking a design minimizing quantities of the type (6.2.9) may arise if for the measure of distinction between the regression curves  $\eta_j(x, \theta_j)$  and  $\eta_k(x, \theta_k)$  the difference

$$Z = S_j(N) - S_k(N)$$

is used, or some sufficiently smooth function  $P_j(Z)$  of this quantity.

Indeed, let the central moments  $d_l(Z)$  ( $l \geq 2$ ) be finite and let the derivatives  $P_j^{(l)}(Z)$ , beginning with the second, be small in the region  $\pm d_l$  ( $l \geq 2$ ), then

$$\begin{aligned} E[P_j(Z)] &= P_j[E(Z)] + P_j[E(Z)] E[Z - E(Z)] \\ &\quad + \frac{1}{2} P_j''[E(Z)] E[Z - E(Z)]^2 + \\ &\simeq P_j[E(Z)], \end{aligned} \quad (6.2.10)$$

and we arrive at (6.2.9).

III. After  $N$  observations are taken at the points  $x_1, x_2, \dots, x_n$ , let a set of measurements be taken at the point  $x$  with weight  $w(x) = \Delta N \lambda(x)$ . We explain how the quantity  $S(N + \Delta N)$  behaves in this situation. For simplicity we assume that  $\eta(x, \theta) = \sum_{a=1}^m \theta_a f_a(x)$ . Generalization to the case of best quasi-linear estimators will be obvious.

**Lemma 6.2.1.** *If at the point  $x$ , a set of observations with weight  $w = \lambda(x) \Delta N$  is taken, then*

$$d(x_1, x_2, N + \Delta N) = d(x_1, x_2, N) - \frac{w d(x_1, x, N) d(x_2, x, N)}{1 + w d(x, N)}, \quad (6.2.11)$$

where  $d(x_i, x_j, N)$  is the covariance between the estimates of the response surface at  $x_i$  and  $x_j$  after  $N$  observations.

*Proof.* Multiplying inequality (4.2.4) on the left by  $f'(x_1)$  and on the right by  $f(x_2)$ , we obtain

$$\begin{aligned} & f'(x_1) D(N + \Delta N) f(x_2) \\ &= f'(x_1) D(N) f(x_2) - \frac{wf'(x_1) D(N) f(x) f'(x) D(N) f(x_2)}{1 + w d(x, N)}. \end{aligned}$$

Considering that  $d(x_i, x_j, N) = f'(x_i) D(N) f(x_j)$  (cf. Section 1.3), it is not difficult to obtain the validity of the lemma being proved.

**Lemma 6.2.2.** *If at the point  $x$ ,  $\Delta N$  observations are taken with the sum of weights  $w = \lambda(x) \Delta N$  and the results  $y_1, y_2, \dots, y_{\Delta N}$ , then*

$$\hat{\eta}(\tilde{x}, N + \Delta N) = \hat{\eta}(\tilde{x}, N) + \frac{w d(\tilde{x}, x, N)[y - \hat{\eta}(x, N)]}{1 + w d(x, N)}, \quad (6.2.12)$$

where  $y = (\Delta N)^{-1} \sum_{j=1}^{\Delta N} y_j$ ,  $\hat{\eta}(x, N)$  is the best linear estimate of the response surface after  $N$  observations, and the remaining notation is the same as in the preceding lemma.

*Remark.* In what follows, for simplicity, we will call the quantity  $y$  the result of observations at the point  $x$ .

*Proof.* By (1.3.6),

$$\hat{\theta} = D(N + \Delta N) Y(N + \Delta N), \quad (6.2.13)$$

where  $D(N + \Delta N)$  is defined by (4.2.4), and

$$Y(N + \Delta N) = \sum_{i=1}^n y_i w_i f(x_i) + w y f(x).$$

Multiplying both sides of (6.2.13) on the left by  $f'(\bar{x})$  and carrying out simple computations, we obtain

$$\begin{aligned}\hat{\eta}(\bar{x}, N + \Delta N) &= f(\bar{x}) \left[ I_m - \frac{w D(N) f(v) f'(x)}{1 + w d(x, N)} \right] D(N) \left[ \sum_{i=1}^n y_i w_i f(x_i) + w y f(x) \right] \\ &= f(v)\theta + \frac{w f(\bar{x}) D(N) f(x)(y - \hat{\eta}(x, N))}{1 + w d(v, N)}\end{aligned}$$

Using the notation introduced above, the last expression can be rewritten in the form (6.2.12). The lemma is proved.

**Lemma 6.2.3 [42]** *Under the assumptions of Lemma 6.2.2 the sum of the weighted square residuals after  $N + \Delta N$  observations can be represented in the form*

$$S(N + \Delta N) = S(N) + \{[y - \hat{\eta}(v, N)]^2 s(x, N)\} \quad (6.2.14)$$

where  $s(v, N) = w^{-1} + d(x, N)$

*Proof* By definition

$$S(N + \Delta N) = \sum_{i=1}^n w [y_i - \hat{\eta}(x_i, N + \Delta N)]^2 + u [y - \hat{\eta}(x, N + \Delta N)]^2$$

Using Lemma 6.2.2 we transform the given equality to the form

$$\begin{aligned}S(N + \Delta N) &= S(N) + \sum_{i=1}^n u \mathcal{A}^*(x_i, x) \\ &\quad - 2 \sum_{i=1}^n w_i [y_i - \hat{\eta}(x_i, N)] \mathcal{A}(x_i, x) \\ &\quad + u \{y - \hat{\eta}(x, N) - \mathcal{A}(v, x)\}^2 \quad (6.2.15)\end{aligned}$$

where

$$\mathcal{A}(x_i, x) = \frac{d(x_i, x, N)[y - \hat{\eta}(v, N)]}{s(x, N)} \quad (6.2.16)$$

Expression (6.2.15) can be simplified if we consider that

$$\begin{aligned} \sum_{i=1}^n w_i d^2(x_i, x, N) &= \sum_{i=1}^n w_i f'(x) D(N) f(x_i) f'(x_i) D(N) f(x) \\ &= f'(x) D(N) M(N) D(N) f(x) = d(x, N). \end{aligned}$$

By analogous operations it is not difficult to verify two further relations:

$$\sum_{i=1}^n w_i y_i d(x_i, x, N) = \hat{\eta}(x, N),$$

$$\sum_{i=1}^n w_i \hat{\eta}(x_i, N) d(x_i, x, N) = \hat{\eta}(x, N).$$

Setting the given identities in (6.2.15), after a simple transformation we obtain (6.2.14). The lemma is proved.

**IV.** We turn to computing quantities of the type (6.2.9) for a given experimental design.

For this it is necessary to know the posterior distribution of the parameters  $\theta$  and the response surface  $\eta(x, \theta)$ .

**Lemma 6.2.4.** *Let measurements be taken at the points  $x_1, x_2, \dots, x_n$  with the results  $y = \|y_1, y_2, \dots, y_n\|$ . Then the conditional density of the sought parameters is equal to*

$$p(\theta | y) = (2\pi)^{-m/2} |D(N)|^{-1/2} \exp[-\frac{1}{2}(\hat{\theta} - \theta)' D^{-1}(N)(\hat{\theta} - \theta)], \quad (6.2.17)$$

and the conditional density of the values of the response surface  $\eta(x)$  at the given point in the factor space is equal to

$$p(\eta(x) | y) = [2\pi d(x, N)]^{-1/2} \exp\left(-\frac{1}{2} \frac{[\eta(x) - \hat{\eta}(x, N)]^2}{d(x, N)}\right), \quad (6.2.18)$$

where  $\hat{\theta}(N)$  is the best linear estimate for the sought parameters,  $D(N)$  is their dispersion matrix, and  $\hat{\eta}(x, N) = f'(x) \hat{\theta}(N)$ .

*Proof.* (1) By Bayes's formula (cf., e.g., [12])

$$p(\theta | y) = \frac{p(y | \theta) p_0(\theta)}{\int p(y | \theta) p_0(\theta) d\theta}, \quad (6.2.19)$$

where  $p(y | \theta)$  is the conditional density of the results of observations for a given  $\theta$ , and  $p_0(\theta)$  is the *a priori* density of the parameters  $\theta$ .

If, *a priori*, all values of  $\theta$  are equally likely, then  $p_0(\theta) \sim \text{const}$  (cf., e.g., [45]) and

$$p(\theta | y) \sim p(y | \theta) \quad (6.2.20)$$

From (6.2.20) and (6.2.1)

$$p(\theta | y) \sim \prod_{i=1}^n \prod_{j=1}^m p(y_i | \eta(x_i, \theta)) \sim \exp \left( -\frac{1}{2} \sum_{i=1}^n w_i [y_i - f(x_i)\theta]^2 \right) \quad (6.2.21)$$

We transform the exponents (for brevity the argument  $N$  is omitted)

$$\begin{aligned} \sum_{i=1}^n w_i [y_i - f(x_i)\theta]^2 &= \sum_{i=1}^n w_i [y_i - f(x_i)\hat{\theta}]^2 \\ &\quad + 2 \sum_{i=1}^n w_i [y_i - f(x_i)\hat{\theta}] [f(x_i)\hat{\theta} - f(x_i)\theta] \\ &\quad + \sum_{i=1}^n w_i [f(x_i)\hat{\theta} - f(x_i)\theta]^2 \end{aligned} \quad (6.2.22)$$

The second term in the right hand side of (6.2.22) in view of the definition of  $\hat{\theta}$  [cf. (1.3.6) and (1.3.9)] is equal to zero

$$\begin{aligned} \sum_{i=1}^n w_i [y_i - f(x_i)\hat{\theta}] [f(x_i)\hat{\theta} - f(x_i)\theta] \\ &= \left[ \sum_{i=1}^n w_i y_i f(x_i) - \hat{\theta} \sum_{i=1}^n w_i f(x_i) f(x_i) \right] (\hat{\theta} - \theta) \\ &= [Y - M\hat{\theta}] (\hat{\theta} - \theta) = 0 \end{aligned} \quad (6.2.23)$$

The last summation in the right-hand side of (6.2.22) can be represented in the form [cf. (1.3.9) and (1.3.10)]

$$\begin{aligned} \sum_{i=1}^n w_i [f(x_i)\hat{\theta} - f(x_i)\theta]^2 &= (\hat{\theta} - \theta) \sum_{i=1}^n w_i f(x_i) f(x_i) (\hat{\theta} - \theta) \\ &= (\hat{\theta} - \theta) D^{-1} (\hat{\theta} - \theta) \end{aligned} \quad (6.2.24)$$

From (6.2.22)–(6.2.24) we have

$$\sum_{i=1}^n w_i [y_i - f'(x_i)\theta]^2 = \sum_{i=1}^n w_i [y_i - \hat{\eta}(x_i)]^2 + (\hat{\theta} - \theta)' D^{-1}(\hat{\theta} - \theta). \quad (6.2.25)$$

Setting (6.2.25) into (6.2.21) and using the well-known fact that

$$\int \exp[-\frac{1}{2}(\hat{\theta} - \theta)' D^{-1}(\hat{\theta} - \theta)] d\theta = (2\pi)^{m/2} |D|^{1/2}, \quad (6.2.26)$$

we obtain (6.2.17).

(2) Since  $\eta(x)$  is a linear combination of the parameters  $\theta$ , the posterior distribution of which is the normal law, then for the determination of the posterior distribution it is sufficient to know its mean and variance. From (6.2.17) and Theorem 1.3.1

$$E[\eta(x)] = \hat{\eta}(x) \quad \text{and} \quad D[\eta(x)] = d(x).$$

It follows that

$$p(\eta(x) | y) = [2\pi d(x, N)]^{-1/2} \exp\left(-\frac{1}{2} \frac{[\eta(x) - \hat{\eta}(x)]^2}{d(x, N)}\right).$$

The lemma is proved.

The results of Lemma 6.2.4 make it possible to compute quantities of the type (6.2.8) for a given experimental design.

We consider the design concentrated at one point.

**Theorem 6.2.1.** [42]. *If at the point  $x$ ,  $\Delta N$  observations are taken with total weight  $w = \lambda(x) \Delta N$ , then*

$$E_j[S_j(N + \Delta N) - S_k(N + \Delta N)]$$

$$= S_j(N) - S_k(N) - \frac{[\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2 + d_j(x, N) - d_k(x, N)}{s_k(x, N)}, \quad (6.2.27)$$

where the operator  $E_j$  indicates expectation with respect to the results of observations at the point  $x$  under the assumption that hypothesis  $H_j$  ( $j = 1, 2, k \neq j$ ) is true, and the remaining notation is the same as in the preceding section.

*Proof.* We find the posterior distribution of the results of observations

at the given point of the factor space under the assumption of the validity of the  $j$ th hypothesis. From (6.2.1) and Lemma 6.2.4

$$\begin{aligned} p_j(y | x, N) &= \int p_j[y | \eta_j(x, \theta_j)] p(\theta_j | y) d\theta_j, \\ &= (2\pi[b^*(x) + d_j(x, N)])^{-1/2} \exp\left(-\frac{1}{2} \frac{[y - \hat{\eta}_j(x, N)]^2}{b^*(x) + d_j(x, N)}\right), \end{aligned} \quad (6.2.28)$$

where  $b_2(x) = \lambda^{-1}(x)$

If at the point  $x$  several observations  $y_1, y_2, \dots, y_{4N}$  are taken, then the mean value of the results  $y = (4N)^{-1} \sum_{j=1}^{4N} y_j$  will be distributed according to the law (6.2.28) with  $b^*(x) = \lambda^{-1}(x)$  replaced by  $b^*(x) = [\lambda(x) 4N]^{-1}$ .

From Lemma 6.2.3,

$$\begin{aligned} S_j(N + \Delta N) - S_k(N + \Delta N) &= S_j(N) - S_k(N) \\ &\quad - \frac{[y - \hat{\eta}_j(x, N)]^2}{s_j(x, N)} + \frac{[y - \hat{\eta}_k(x, N)]^2}{s_k(x, N)} \end{aligned} \quad (6.2.29)$$

Taking expectation of both sides of (6.2.29) we obtain from the computation (6.2.28) that

$$\begin{aligned} L_j[S_j(N + \Delta N) - S_k(N + \Delta N)] &= S_j(N) - S_k(N) \\ &\quad - \frac{s_j(x, N) + [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2}{s_k(x, N)} + 1 \\ &= S_j(N) - S_k(N) \\ &\quad - \frac{[\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2 + d_j(x, N) - d_k(x, N)}{s_k(x, N)}, \end{aligned}$$

which is what was required to be shown.

From the theorem just proved it follows that if restricted to the collection of one-point designs, then the optimal design  $\delta^*(\Delta N)$  consists of  $\Delta N$  observations taken at the point corresponding to

$$\min_x \sum_{\substack{j=1 \\ j \neq k}}^2 p_j(H_j) \gamma_k P_j(E_j[S_j(N_0 + \Delta N) - S_k(N_0 + \Delta N)]) \quad (6.2.30)$$

where

$$\begin{aligned} E_j[S_j(N_0 + \Delta N) - S_k(N_0 + \Delta N)] \\ = S_j(N_0) - S_k(N_0) \\ - \frac{[\hat{\eta}_j(x, N_0) - \hat{\eta}_k(x, N_0)]^2 + d_j(x, N_0) - d_k(x, N_0)}{s_k(x, N_0)}. \end{aligned}$$

In some cases the aim of conducting experiments is not defined in a form which permits one to find the relationship between the loss from incorrectly accepting the first model and the loss from incorrectly accepting the second model. In such situations each  $(N + 1)$ st observation must be taken at the point corresponding to

$$\max_x \min_{\substack{j, k=1, 2 \\ k \neq j}} E_k[S_j(N + 1) - S_k(N + 1)]. \quad (6.2.31)$$

It is evident that the construction of the design  $\mathcal{E}(\Delta N)$  is possible if a preliminary experiment is conducted according to some design  $\mathcal{E}(N_0)$  which is nondegenerate with respect to the regression curves  $\eta_1(x, \theta_1)$  and  $\eta_2(x, \theta_2)$ , that is, the spectrum must consist of at least  $m = \max(m_1, m_2)$  points, where  $m_1$  and  $m_2$  respectively are the dimensions of the vectors  $\theta_1$  and  $\theta_2$ .

From (6.2.30) it follows that each  $N$ th point is chosen in such a way that as a result of the  $N$ th observation [for the form of  $\eta_1(x, \theta_1)$  and  $\eta_2(x, \theta_2)$  given earlier], the averaged possible losses for accepting an incorrect hypothesis  $l(N)$  [cf. (6.2.9)] decreases as rapidly as possible.

V. The position of the point  $x_N$  changes depending on the value of  $N$ . This is explained by the fact that the rate of decrease of the quantity  $l(N, x)$  falls with the growth of  $N$  when taking observations at one point of the factor space. By a direct differentiation of  $l(N, x)$  it is easy to verify that

$$(\partial/\partial N) | l(N, x) | \sim O[(\Delta N)^{-2}], \quad (6.2.32)$$

where  $\Delta N$  is the number of observations taken at the point  $x$ . In (6.2.32) it was assumed that the function  $p_j(z)$  is close to linear in the neighborhood of  $z$  under study. In the contrary case the dependence of the derivative  $\partial/\partial N | l(\Delta N, x) |$  on  $\Delta N$  changes by comparison with (6.2.32), but the character of  $l(N)$  is asymptotically preserved.

We pursue more detailed reasons for the movement of the point  $x_N$  in the factor space as  $N$  grows. We will assume that the following inequalities are satisfied

$$|d_j(x, \tilde{x}, N)| \ll [\lambda(x) \lambda(\tilde{x})]^{-1/2}, \quad (6.2.33)$$

and

$$|d_j(x, \tilde{x}, N)| \ll [|\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)| |\hat{\eta}_j(\tilde{x}, N) - \hat{\eta}_k(\tilde{x}, N)|], \quad (6.2.34)$$

where  $j = 1, 2, j \neq k$

Inequalities (6.2.33) and (6.2.34) usually hold for  $N \gg m_j$  ( $m_j$  is the number of unknown parameters for the hypothesis  $H_j$ ). For example, for continuous  $D$ -optimal designs  $\mathcal{E}_j(N)$  (cf. Theorem 2.2.2),

$$\max_x \lambda(x) d_j(x, N) = m_j N^{-1},$$

and it follows that

$$d_j(x, N) \leq \lambda^{-1}(x) m_j N^{-1} \quad (6.2.35)$$

From (6.2.35) and the inequality

$$\left| \frac{d(x, \tilde{x}, N)}{[d(x, N) d(\tilde{x}, N)]^{1/2}} \right| \leq 1,$$

and from the definition of the covariance  $d(x, \tilde{x}, N)$ , it follows that

$$|d_j(x, \tilde{x}, N)| \leq m_j N^{-1} [\lambda(x) \lambda(\tilde{x})]^{-1/2}$$

For an arbitrary nondegenerate design  $\mathcal{E}(N)$ , distinct from the  $D$ -optimal one, the last inequality passes to a weaker condition

$$|d_j(x, \tilde{x}, N)| = O(m_j N^{-1} [\lambda(x) \lambda(\tilde{x})]^{1/2}) \quad (6.2.36)$$

In this manner, to satisfy inequality (6.2.33) it is sufficient that  $m_j N^{-1} \ll 1$  ( $j = 1, 2$ ). Inequality (6.2.34) for the region where  $\hat{\eta}_j(x, N)$  and  $\hat{\eta}_k(x, N)$  are distinct from one another is a corollary of the first inequality.

Let an observation be taken at the point  $x_N$  with weight  $w(x_N)$ . Then from Lemmas 6.2.1 and 6.2.2 we obtain ( $j = 1, 2$ )

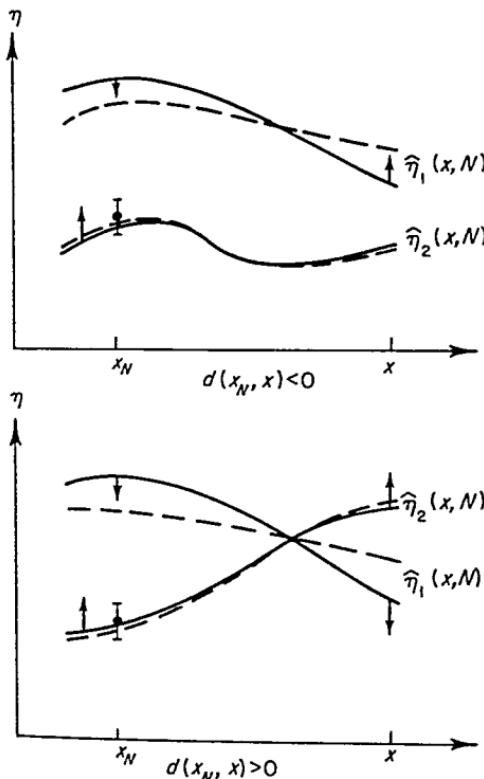
$$\hat{\eta}_j(x, N+1) = \hat{\eta}_j(x, N) + \frac{w(x_N) d_j(x, x_N, N) [y - \hat{\eta}_j(x_N, N)]}{1 + w(x_N) d_j(x_N, N)},$$

$$d_j(x, x_N, N+1) = d_j(x, x_N, N) - \frac{w(x_N) d_j^2(x, x_N, N)}{1 + w(x_N) d_j(x_N, N)},$$

or, keeping terms of the first order in  $d_j(x, x_N, N)$ ,

$$\begin{aligned}\hat{\eta}_j(x, N+1) - \hat{\eta}_k(x, N+1) &= \hat{\eta}_j(x, N) - \hat{\eta}_k(x, N) \\ &\quad - w(x_N)\{d_k(x, x_N, N)[y - \hat{\eta}_k(x_N, N)] \\ &\quad - d_j(x, x_N, N)[y - \hat{\eta}_j(x_N, N)]\}, \\ d_j(x, x_N, N+1) &= d_j(x, x_N, N)[1 - w(x_N)d_j(x_N, N)].\end{aligned}\tag{6.2.37}$$

From (6.2.37) it follows (cf. also Fig. 25) that the curves  $\hat{\eta}_1(x)$  and



**Fig. 25.** Variation of the estimate of the regression curves  $\eta_1(x)$  and  $\eta_2(x)$  in conducting complimentary observations. Results of observation are plotted by points; half of the length of the interval, the center of which is the specified point, is equal to the error of observation. The solid line corresponds to the estimates of the curves  $\eta_j(x)$  up to conducting supplementary measurements, the dashed line to the estimates after taking supplementary measurements.

$\hat{\eta}_2(x)$  have a tendency to be closer at the point  $x_N$  and, depending on the sign of  $d_j(x, x_N, N)$ ,  $j = 1, 2$ , to be closer or further at other points of the factor space

In order to prove this, it is sufficient to take the expectation on both sides of (6.2.37). Then, under the assumption that the  $j$ th hypothesis is true,

$$\begin{aligned} L_j[\hat{\eta}_j(x, N+1) - \hat{\eta}_k(x, N+1)] \\ = \hat{\eta}_j(x, N) - \hat{\eta}_k(x, N) - w(x_N) d_k(x, x_N, N) [\hat{\eta}_j(x_N, N) - \hat{\eta}_k(x_N, N)] \end{aligned} \quad (6.2.38)$$

Since  $d_k(x_V, x_N, N) = d_k(x_N, N)$ , then

$$|L_j[\hat{\eta}(x_N, N+1) - \hat{\eta}_k(x_N, N+1)]| < |\hat{\eta}(x_N, N) - \hat{\eta}_k(x_N, N)|$$

At the remaining points the sign of the corrections are determined by the sign of the covariances  $d_k(x, x_N, N)$  (cf Fig 25)

Approaching the middle between the curves  $\hat{\eta}_1(x_N)$  and  $\hat{\eta}_2(x_V)$  brings out the worst condition of accumulation of information at the point  $x_N$  necessary for discrimination of the hypothesis

In connection with this, it becomes obvious that for large  $\Delta N$  the one-point design  $\delta(\Delta N)$  becomes less useful than the design consisting of several points

The construction of optimal discriminating designs, which minimize quantities of the type (6.2.9) and have spectrums consisting of several points, is a much more complicated computational problem. In this regard methods of discriminating hypotheses relying on a fixed sample size, for a given experimental design, require on the average significantly more observations for the achievement of a given accuracy than sequential criteria [49-51].

Therefore it is natural to turn to sequential methods of discriminating hypothesis and to sequential designs of discriminating experiments. Moreover, as we will see later, sequential designs are simple from the computational point of view and they also permit the use of information acquired as a result of the observations taken.

Here it will be appropriate to carry out the analog of the design of experiments in the determination of estimates of several parameters for nonlinear parametrization.

As was shown in Chapter 2, it is possible in this case to construct only locally optimal designs  $\delta(N, \theta_0)$ . The characteristics of these

designs depend critically on the true values of the parameters, which naturally are unknown before the experiment. In connection with this possibility, the sequential construction (cf. Chapter 3) of only the locally  $D$ -optimal designs is indicated. The essence of this design consists of the following. The total number of observations, used in the entire experiment, is divided into small parts. Observations are then conducted at points which are the most optimal from the point of view of acquiring knowledge at the given moment [or more precisely, the points of the factor space at which the next group of observations should be taken as determined by the values  $\hat{\theta}(N)$  and  $D(N)$ ].

In this situation the closer the estimates  $\hat{\theta}(N)$  are to the true values of the parameters  $\theta_0$  the closer the allocation of observations in the factor space becomes to the locally optimal design. The situation for the discrimination of hypotheses is also very complicated.

For example, let the true model be the first model  $\eta_1(x, \theta_1)$  and the true value of the parameter be  $\theta_{10}$ . In this case, locally optimal designs maximize the difference

$$\Delta = S_2(N) - S_1(N, \theta_0),$$

or, what is the same thing, some monotone-increasing function of  $\Delta$  (cf. Part II of the current section).

Since the true model and the true value of the corresponding sought parameters are unknown, we find it necessary to act as if there were a single model. The total number of possible observations is divided into small parts. After carrying out the grouping of the observations an analysis of the obtained experimental data is conducted. Relying on the obtained reduction we find the optimal (from the point of view of already available information) distribution of subsequent observations or groups of observations, etc. Turning to somewhat looser terminology, the idea of sequential design in the given case can be formulated in the following manner:

Reconnaissance of nature—design—  
reconnaissance of nature—design—... .

With each new stage, the reconnaissance of nature becomes more purposeful.

**VI.** We assume now that the hypotheses are tested according to the sequential design which is defined by means of the following decision rule.

1 Hypothesis  $H_1$  is accepted if

$$P[S_1(N) - S_2(N)] \leq A$$

2 Hypothesis  $H_2$  is accepted if

$$P[S_2(N) - S_1(N)] \leq B$$

3 Observations are continued if

$$A \leq P[S_1(N) - S_2(N)] \quad \text{or} \quad B \leq P[S_2(N) - S_1(N)]$$

The function  $P(Z)$  must be such that for any  $Z$ , one of the three inequalities just introduced can hold. Sequential criteria [49, 50] using a suitable function  $P(Z)$  permits a significant decrease in the average number of observations necessary for attaining a given value of accuracy in discriminating hypotheses.

We consider an experiment conducted according to the following strategy:

1 After the  $N$ th observation the quantities  $d_j(x, N)$ ,  $\eta_j(x, N)$  ( $j = 1, 2$ ) are computed

2 The point  $x_N$  is found corresponding to

$$\min_x \{W_1(N) E_1[S_1(N) - S_2(N)] + W_2(N) E_2[S_2(N) - S_1(N)]\} \quad (6.2.39)$$

where

$$E_j[S_j(N+1) - S_k(N+1)]$$

$$= S_j(N) - S_k(N) - \frac{[\hat{\eta}_j(x, N) - \eta_k(x, N)]^2 + d_j(x, N) - d_k(x, N)}{s_k(x, N)}$$

The weight function  $W_j(N)$  ( $j = 1, 2$ ) characterizes the degree of faith the experimenter has in the  $j$ th hypothesis after  $N$  observations. The choice of these functions is defined by the form of the functions  $P[S_j - S_k]$ , the cost  $\gamma_j$ , and the *a priori* probability  $p_0(H_j)$  ( $j = 1, 2$ ) (cf. the example at the end of this section).

3 An observation is taken at the point  $x_N$ .

4 Operations 2 and 3 are repeated with the index  $N$  changed to  $N + 1$ , and so on as long as all means allocated to the experiment are not exhausted or the possible loss  $l(N)$  does not drop below a given level. On computing the quantities  $d(x, N)$ ,  $\eta_j(x, N)$  ( $j = 1, 2$ ) after

each supplementary observation, it is useful to use the results of Lemmas 6.2.1 and 6.2.2. The sequential procedure indicates the point where the observation will be taken at a given moment (after the  $N$ th observation) so that the most information is obtained from the point of view of discriminating hypotheses [in the decision rule, relying on the difference  $S_2(N) - S_1(N)$ ]. It is evident in this case that the average number of observations necessary for discriminating the hypotheses is essentially reduced in comparison with the non-sequential experimental design.

**VII.** In conclusion, we note that from the methods of constructing the best linear estimate (cf. Section 1.3) and the definition of the decision rule, it follows that the discrimination of hypotheses about the form of the response surface is possible with any given accuracy (the loss for accepting the false hypotheses can be made arbitrarily small) when a distribution of allocation  $\xi_j(x)$  [cf. Section 1.10,  $\int_X d\xi_j(x) = 1$ ], satisfying the requirement

$$\int_X \lambda(x)[\hat{\theta}_j f_j(x) - \hat{\theta}_k f_k(x)]^2 d\xi > 0, \quad (6.2.40)$$

can be found for the space  $X$  of possible observations and the true hypotheses  $H_j$ .

The quantity  $\hat{\theta}_k$  is defined by

$$\hat{\theta}_k = M_k^{-1} Y_k,$$

where

$$M_k^{-1} = \int_X \lambda(x) f_k(x) f_k'(x) d\xi_j(x), \quad Y_k = \int_X \lambda(x) \eta_j(x) f_k(x) d\xi_j(x).$$

Inequality (6.2.40) says that we can as a result of some experiment obtain the validity of the hypothesis  $H_j$ , when the true response surface is not approximated sufficiently accurately [in the sense of the metric (6.2.40)] by the curve  $\hat{\eta}_k(x, \theta_k) = \hat{\theta}_k' f_k(x)$  ( $k \neq j$ ).

**EXAMPLE.** Let some process be described by one of the two functions

$$\eta(x, \theta) = \begin{cases} \eta_1(x, \theta_1) = \theta x & (\text{hypothesis } H_1); \\ \eta_2(x, \theta_2) = \phi x^2 & (\text{hypothesis } H_2). \end{cases}$$

Based on experiments which measure the quantity  $y$  it is necessary to

clarify which of the two models is valid (i.e., obtain for the false model sufficiently large values of the sum of the weighted quadratic deviations)

The experiment, measuring the quantity  $y$ , is characterized by the efficiency

$$\lambda(x) = x, \quad 0 \leq x \leq 1$$

Preliminary observations are conducted at the point  $x = 1, w = 2$ . The given design is  $D$ -optimal for both hypotheses. The average of the observations for the two models equals 1.1. For the unknown parameters the following estimates are obtained

$$\theta = 1.1, \quad D(\theta) = 0.71, \quad \phi = 1.1, \quad D(\phi) = 0.71$$

As a function characterizing the accuracy of the discrimination of the hypothesis we choose the difference of the sums of the weighted quadratic deviations

$$P[S_1(N) - S_2(N)] = S_1(N) - S_2(N)$$

We will accept the first hypothesis if

$$S_1(N) - S_2(N) \leq A$$

the second, if

$$S_2(N) - S_1(N) \leq A$$

and continue taking observations if

$$|S_1(N) - S_2(N)| < A,$$

where  $A$  is a constant given before the experiment. If the results of the observations are distributed according to a normal law and both hypotheses are *a priori* equally likely, then it is natural to choose the weight entering in (6.2.39) equal to

$$W_j(N) \sim e^{-S_j(N)/2}$$

In Fig. 26 we present the position of the optimal point, sought in accordance with the iterative procedure 1-4 under the condition that at each step the following minimum is found

$$\min_x (e^{-S_1(N)/2} E_1[S_1(N) - S_2(N)] + e^{-S_2(N)/2} E_2[S_2(N) - S_1(N)])$$

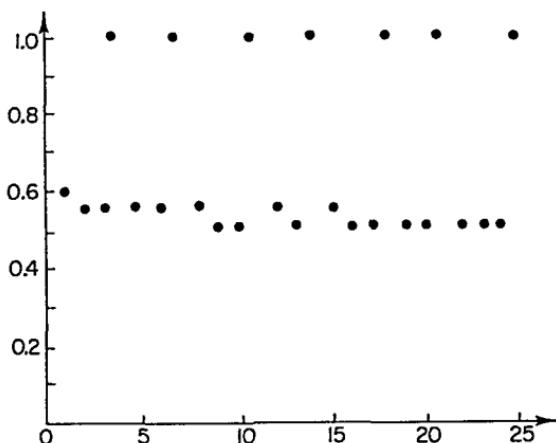


Fig. 26. Position of the optimal points of observation for the discrimination of the models  $\theta_x$  and  $\phi x^2$ .

The results are simulated with a table of random numbers under the assumption that the true hypothesis is  $H_2$  ( $\phi = 1$ ).

For obtaining the boundary  $A = 1$ , in this case 47 observations were required. For obtaining the boundary with a distribution of observations uniform on ten equidistant points  $\sim 120$  simulated observations were required.

### 6.3. The Method of Likelihood Ratio

Let the hypotheses  $H_1$  and  $H_2$  be formulated as in Section 6.1 and let the results of observations be independent and normally distributed.

We choose as a measure of the estimate of accuracy of the discriminating experiment, the modulus of the logarithm of the generalized likelihood ratio which, for *a priori* equally likely hypotheses, has the form

$$\frac{L_1(N)}{L_2(N)} = \frac{\int p_1[y_1 | \eta_1(x_1, \theta_1)] p_1[y_2 | \eta_1(x_2, \theta_1)] \cdots p_1[y_N | \eta_1(x_N, \theta_1)] d\theta_1}{\int p_2[y_1 | \eta_2(x_1, \theta_2)] p_2[y_2 | \eta_2(x_2, \theta_2)] \cdots p_2[y_N | \eta_2(x_N, \theta_2)] d\theta_2}. \quad (6.3.1)$$

Just as before, we will assume that  $\eta_j(x) = f'_j(x) \theta_j$  ( $j = 1, 2$ ). Generalization to the case of nonlinear parametrization is conducted by the

method of quasi linearization and can be carried out by the reader independently

**Lemma 6.3 1.** *Under the assumption of normality of the conditional distribution  $p_j[y | \eta_j(x, \theta_j)]$ , the generalized likelihood function is equal to*

$$L_j(N) = (2\pi)^{(m_j-N)/2} \prod_{i=1}^N w_i^{1/2} |D_j(N)|^{1/2} e^{-S_j(N)/2}, \quad (6.3.2)$$

where  $m_j$  is the number of unknown parameters under the  $j$ th hypotheses,  $w_i = b_i^{-2}$ ,  $S_j(N) = \sum_{i=1}^N w_i [y_i - f_j'(x_i) \theta_j]^2$ , and  $D_j(N) = M_j^{-1}(N)$  is the dispersion matrix of the best linear estimate  $\hat{\theta}_j$ .

*Proof* For brevity, the index  $j$  in the proof will be omitted. From the definition of  $L(N)$  it follows that

$$L(N) = \int (2\pi)^{-N/2} \prod_{i=1}^N w_i^{1/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^N w_i [y_i - f(x_i)\theta]^2\right\} d\theta \quad (6.3.3)$$

As was shown in the proof of Lemma 6.2.4,

$$\begin{aligned} \sum_{i=1}^N w_i [y_i - f(x_i)\theta]^2 &= \sum_{i=1}^N w_i [y_i - f(x_i)\hat{\theta}]^2 \\ &\quad + (\hat{\theta} - \theta) D^{-1}(N)(\hat{\theta} - \theta) \end{aligned} \quad (6.3.4)$$

Setting (6.3.4) into (6.3.3) and integrating with respect to  $\theta$  [cf (6.2.26)] we obtain (6.3.2). The lemma is proved.

From Lemma 6.3.1 it follows that

$$\frac{L_1(N)}{L_2(N)} = (2\pi)^{(m_1-m_2)/2} \frac{|D_1(N)|^{1/2}}{|D_2(N)|^{1/2}} \exp\{\frac{1}{2}[S_2(N) - S_1(N)]\} \quad (6.3.5)$$

We note that (6.3.5) does not change its value if the results of the observations  $y_{i1}, y_{i2}, \dots, y_{in_i}$  are grouped and combined into one point and we replace  $n_i$  terms occurring in  $S_i(N)$  by one term  $w_i[y_i - f'(x_i)\hat{\theta}]^2$ , where

$$y_i = n^{-1} \sum_{j=1}^n y_{ij} \quad w_i = n b_j^{-2} = n_i \lambda(x_i)$$

By assumption  $|\ln[L_1(N)/L_2(N)]|$  is a measure of accuracy of the

discriminating experiment; therefore the experimenter must place his observations in the factor space in such a way that the fastest growth of  $|\ln[L_1(N)/L_2(N)]|$  as a function of  $N$  is attained.

Suppose  $N$  observations have been taken and the experimenter must choose the point  $x_N$  where the growth  $\Delta |\ln[L_1(N)/L_2(N)]|$  will be maximal. It is not difficult to see that the growth of this increment depends not only on the coordinates of the points where the observation will be taken but also on the result of the observation  $y$ . Therefore, if the  $j$ th hypothesis is true, we maximize the expected value, taken with respect to  $y$ , of the increment [52]:

$$E_j \left[ \ln \frac{L_j(N+1)}{L_k(N+1)} - \ln \frac{L_j(N)}{L_k(N)} \right] \quad (j = 1, 2, \quad k \neq j). \quad (6.3.6)$$

Since *a priori* it is unknown which of the hypothesis is true, then it is necessary in (6.3.6) to average with respect to the probability  $p_N(H_j)$  of each of the hypotheses  $H_j$ , and maximize the expression

$$\gamma(x, N) = \sum_{j=1}^2 p_N(H_j) E_j \left[ \ln \frac{L_j(N+1)}{L_k(N+1)} - \ln \frac{L_j(N)}{L_k(N)} \right], \quad k \neq j. \quad (6.3.7)$$

**Theorem 6.3.1.** *The average growth of the logarithm of the likelihood ratio after conducting an observation at the point  $x$  under the assumption that the  $j$ th hypothesis is true equals*

$$\begin{aligned} & E_j \left[ \ln \frac{L_j(N+1)}{L_k(N+1)} - \ln \frac{L_j(N)}{L_k(N)} \right] \\ &= \frac{1}{2} \ln \frac{s_j(x, N)}{s_k(x, N)} + \frac{1}{2} \frac{[\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2 + d_j(x, N) - d_k(x, N)}{s_k(x, N)}, \end{aligned} \quad (6.3.8)$$

where

$$d_j(x, N) = d[\hat{\eta}_j(x, N)] \quad \text{and} \quad s_j(x, N) = \lambda^{-1}(x) + d_j(x, N).$$

*Proof.* From (6.3.5)

$$\begin{aligned} & \ln \frac{L_j(N+1)}{L_k(N+1)} - \ln \frac{L_j(N)}{L_k(N)} \\ &= \frac{1}{2} \ln \left[ \frac{|D_j(N+1)| |D_k(N)|}{|D_k(N+1)| |D_j(N)|} \right] \\ &+ \frac{1}{2} [S_k(N+1) - S_k(N) - S_j(N+1) + S_j(N)]. \end{aligned} \quad (6.3.9)$$

The first term does not depend on  $y$  and equals [cf. Lemma 4.2.2 or Eq. (4.2.9)]

$$\frac{1}{2} \ln \frac{|D_k(N+1)| |D_k(N)|}{|D_k(N+1)| |D_k(N)|} = \frac{1}{2} \ln \frac{1 + \lambda(x) d_k(x, N)}{1 + \lambda(x) d_k(x, N)} \quad (6.3.10)$$

By Theorem 6.2.1, the expected value of the second term equals

$$\frac{1}{2} \frac{[\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2 + d_j(x, N) - d_k(x, N)}{s_k(x, N)} \quad (6.3.11)$$

Combining (6.3.9)–(6.3.11) we obtain (6.3.8). The theorem is proved.

In those cases where the preliminary experiment conducts not one but  $n$  observations at the point  $x$ , then instead of the efficiency function  $\lambda(x)$  it is necessary to enter the product  $n\lambda(x)$  in formula (6.3.8) [cf. the remarks to (6.3.5)].

Relying on (6.3.7) and Theorem 6.3.1, the following strategy for sequential design may be recommended:

- 1 After the  $N$ th observation the quantities  $\theta(N)$ ,  $\hat{\eta}_j(x, N)$ ,  $d_j(x, N)$ ,  $L_j(N)$  ( $j = 1, 2$ ) are computed
- 2 The point  $x_N$  corresponding to

$$\max_x \sum_{j=1}^2 p_N(H_j) E_j \left[ \ln \frac{L_j(N+1)}{L_k(N+1)} - \ln \frac{L_j(N)}{L_k(N)} \right] \quad (6.3.12)$$

is found. We note that  $p_N(H_j) \sim L_j(N)$ , if the set of hypotheses  $H_j$  ( $j = 1, 2$ ) is complete.

- 3 An observation is taken at the point  $x_N$

Operations 1–3 are then repeated with the index  $N$  replaced by  $N+1$ , and so on as long as the resources have not been exhausted (in this case the  $j$ th hypothesis is accepted if  $L_j/L_k > 1$ ) or  $|\ln[L_1(N)/L_2(N)]|$  does not exceed any of the given boundaries.

It is evident that procedures 1–4 must be preceded by a "priming" experiment conducted according to a nondegenerate design with respect to the regression curves.

Suppose the number of observations allocated to the experiment is not given beforehand and the choice of hypotheses is made according to the decision rule

1. Accept the hypothesis  $H_1$  if

$$\ln L_1(N)/L_2(N) \geq A. \quad (6.3.13)$$

2. Accept the hypothesis  $H_2$  if

$$\ln L_1(N)/L_2(N) \leq B. \quad (6.3.14)$$

3. Continue sampling if

$$B < \ln L_1(N)/L_2(N) < A. \quad (6.3.15)$$

Then the method of design presented maximizes the probability of terminating the process of discriminating hypotheses after the  $(N + 1)$ st observation [for a given  $\hat{\theta}_j(N)$ ,  $d_j(x, N)$ ,  $L_j(N)$ ].

If the cost of observation  $c$  is small, then  $A$  and  $B$  are recommended to be chosen equal to  $\ln c^{-1}$  and  $\ln c$ , respectively.

In this case [53, 54] the decision rule  $\delta_c$ , defined by (6.3.13)–(6.3.15), will be asymptotically optimal for any given experimental design  $\epsilon$ , that is,

$$\lim_{c \rightarrow 0} [r(\delta_c)/r(\hat{\delta})] = 1, \quad (6.3.16)$$

where  $r(\delta)$  is the Bayes risk for the decision rule  $\delta$ , and  $\hat{\delta}$  is the optimal decision rule minimizing the Bayes risk.

In those cases when it is inexpedient (for example, from economic considerations) to conduct an analysis after each distinct observation, the sequential procedure can be modified by allocating at the points  $x_N$  several observations. The quantity  $\lambda(x)$  in (6.3.12) is then replaced by  $\Delta N \lambda(x)$ .

**EXAMPLE.** As a result of analysis of experimental data for some quantity (depolarization  $D_{np}$  for a neutron–proton scatter at an energy of 660 MeV [43]), two possible dependencies on  $x$  were predicted:  $\hat{\eta}_1(x)$  with a standard deviation  $d_1^{1/2}(x)$  and  $\hat{\eta}_2(x)$  with a standard deviation  $d_2^{1/2}(x)$ . Moreover,  $S_1(0) \simeq S_2(0)$  and  $|D_1(0)| = |D_2(0)|$ ; in other words, the true curve  $\eta(x)$  can equally likely belong to the set of curves characterized by either the quantities  $\hat{\eta}_1(x)$ ,  $d_1(x)$  (hypothesis  $H_1$ ) or the quantities  $\hat{\eta}_2(x)$ ,  $d_2(x)$  (hypothesis  $H_2$ ). The curves  $\hat{\eta}_1(x)$  and  $\hat{\eta}_2(x)$  are represented in Fig. 27. The broken lines correspond to the standard deviations  $d_1^{1/2}(x)$  and  $d_2^{1/2}(x)$ . The physical interpretation of the quantity  $\eta(x)$  and the techniques of its measurement are presented in [43].

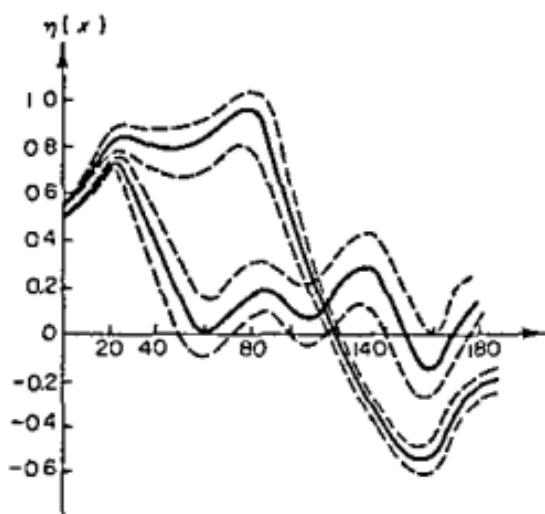


Fig. 27 Predicted value for the quantity  $D_{\alpha\beta}(x) - \eta(x)$

In [43] the efficiency of the experiment in measuring the quantity  $\eta(x)$  is computed. The results are presented in Fig. 28.

A particular characteristic of experiments in the scattering of elementary particles is that a large number of observations are recorded at one time (for a given experiment this number is  $\sim 10^3$  sec $^{-1}$ ). Therefore instead of the number of observations  $N$  it is useful to deal with the time  $T$  necessary for taking the observations. The efficiency function is expressed in corresponding units.

If the mean of the results  $y_i = n_1^{-1} \sum_{j=1}^n y_{ij}$  is already included in the analysis, then for the given conditions the distribution of  $y$  will be

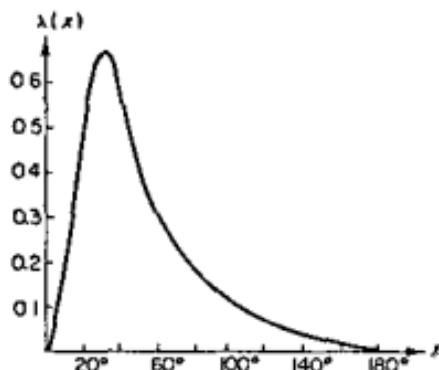


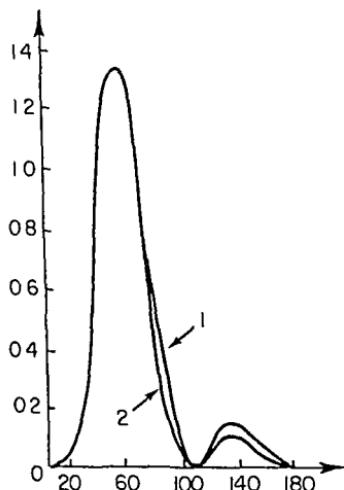
Fig. 28 Efficiency for the experiment in measuring  $D_{\alpha\beta}$

close to the normal and the use of the method developed in this section will be valid.

Suppose it is necessary to design an experiment for measuring quantities which would permit, in a given time, the most effective determination [in the sense of (6.3.1)] of which of the two collections of curves specified above the true curve belongs.

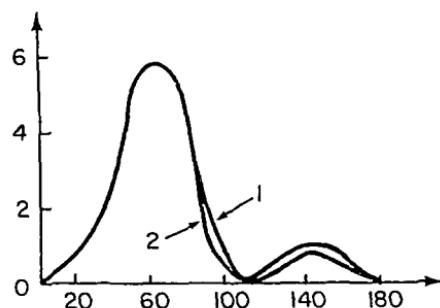
We assume that during the time  $T$ , allotted directly to observations, it is not advantageous to change the apparatus with which the observations of the quantity  $\eta(x)$  are being taken (e.g., the time for setting up and adjusting the apparatus is of the same order as  $T$ ). In this case the design is conducted according to formula (6.3.12) with  $w(x) = \lambda(x)T$ .

In Figs. 29 and 30 the results of designing an experiment in measuring  $\eta(x)$  for  $T = 10$  and  $T = 100$  hr are presented.



**Fig. 29.**  $E_1 \ln(L_1/L_2)$  (curve 1) and  $E_2 \ln(L_1/L_2)$  (curve 2) for  $T = 10$  hr.

**Fig. 30.**  $E_1 \ln(L_1/L_2)$  (curve 1) and  $E_2 \ln(L_1/L_2)$  (curve 2) for  $T = 100$  hr.



A design was also conducted for measuring  $\eta(x)$  which assumed the possibility of changing the measuring apparatus but excluded the averaging reduction of the data

Since the quantities  $d_1(x)$  and  $d_2(x)$  are small (cf. Fig. 27) then by repeating the arguments of Section 6.2, Part IV, it is possible to formulate a rule specifying when to stop taking observations at the point  $x_N$  and start taking observations at the point  $x_{N+1}$ , and so on. Relying on (6.3.7) it is not difficult to see that the observations at the point  $x_N$  must cease as soon as the point  $x_{N+1}$  is found such that

$$\partial\gamma(x_N, t)/\partial t|_{t=t_1} \leq \partial\gamma(x_{N+1}, t)/\partial t|_{t=t_1} \quad (6.3.17)$$

The results obtained on the basis of (6.3.17) for  $T = 100$  are presented in Fig. 31. As a result of the experiment the increment of the quantity  $\gamma(x, T)$  will be equal to 72 that is, 12 times larger (cf. Fig. 30) than in the case of fixed apparatus (this corresponds to the fact that the ratio of probabilities will be three times larger than for the fixed apparatus)

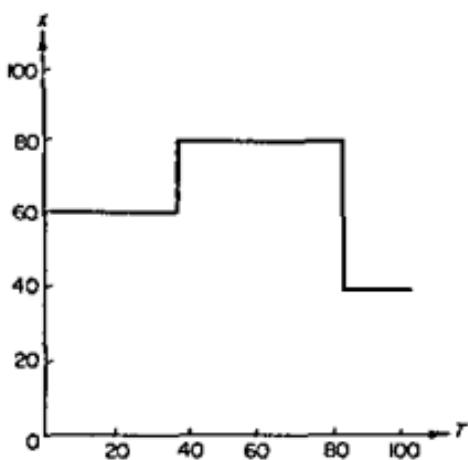


Fig. 31 Optimal position of measuring apparatus as a function of time

For comparing the effectiveness of the design methods presented with the usual intuitive approach (uniform allocation in all regions where observations are practically possible) the quantity  $E \ln(L_1/L_2)$  was computed by formula (6.3.5) for the case when the observations

are taken at the points  $20^\circ, 40^\circ, \dots, 160^\circ$  each for a duration of time  $t_i = 1.25$  hr ( $\sum_{i=1}^8 t_i = 10$  hr).

The computations show that as a result of such observations  $E \ln(L_1/L_2) = 0.33$ . This corresponds to the fact that the probability ratio will be approximately 2.7 times smaller (cf. Fig. 30) than the results of the experiment designed by (6.3.7). We note that the value  $E \ln(L_1/L_2) = 0.33$ , for the allocation of resources corresponding to (6.3.7), is attained in 2.5 hr.

#### 6.4. Discriminating Hypotheses Based on the Entropy Measure of Information

The concept of entropy as a measure of the orderliness of a system is applied in many regions of science. In physics, this measure is the basis for understanding the formulation of many physical laws. The success of using the entropy measure of information in the theory of communications is generally known.

If a system can be found in one of  $v$  states each with probability  $p_j$  ( $j = 1, 2, \dots, v$ ), then the entropy of the system is equal to

$$I = - \sum_{j=1}^v p_j \ln p_j. \quad (6.4.1)$$

In those cases where the system can take on an infinite number of states, determined by some parameter  $t$ , the sum in (6.4.1) is replaced by the integral

$$I = - \int p(t) \ln p(t) dt. \quad (6.4.2)$$

In physical applications the entropy is a measure of the orderliness of a physical system. It is easy to verify that the entropy is maximized for  $p_1 = p_2 = \dots = p_v = v^{-1}$ , i.e., when all states are equally likely (there is no order). The entropy decreases with an increase in the probability of finding the system in one (or several) of the states. As an illustration of this, the value of the entropy, computed for various probability distributions, is presented in Figs. 32A and 32B.

We turn to the problem of discriminating hypotheses. Let there be a system of  $v$  hypotheses and an *a priori* probability of the  $j$ th hypothesis being true equal to  $p_0(H_j)$ .

We assume that an experiment has been conducted. From its

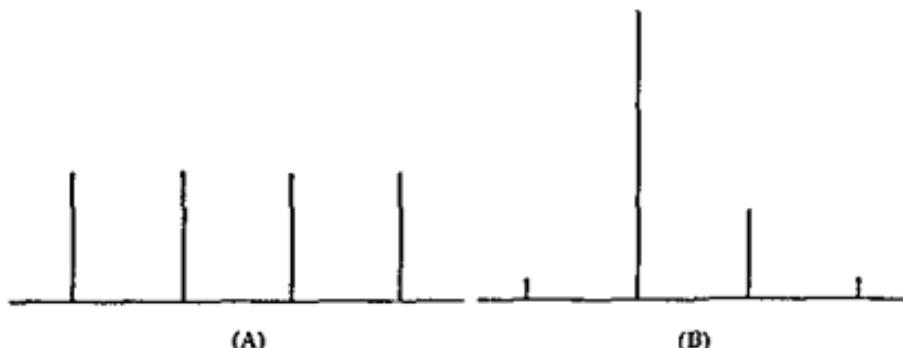


Fig. 31A All four states of the system are equiprobable  $I \approx 1.38$

Fig. 32B The second state of the system dominates  $I \approx 0.51$

results the *a posteriori* probability  $p(H_j)$  is computed. We consider the quantity

$$\Delta I = - \sum_{j=1}^v p_0(H_j) \ln p_0(H_j) + \sum_{j=1}^v p(H_j) \ln p(H_j) \quad (6.4.3)$$

It is evident that  $\Delta I(\mathcal{E})$  can be treated as an increment of information about the system of concurring hypotheses. In the future the quantity  $\Delta I(\mathcal{E})$  will be called the entropy measure of information (or simply the information) obtained in the experiment  $\mathcal{E}$ .

If the accuracy of the experiment (and correspondingly the possible loss) is characterized by the entropy measure of information, then it is natural to require that the design of the experiment  $\mathcal{E}$  be chosen in such a way that the increment of information will be maximal.

We compute the increment of information for the case where experiments are conducted to determine the best mathematical model. Let the observations be taken at points of the factor space  $x_1, x_2, \dots, x_N$ . Then under the assumption that the collection of hypotheses  $H_j$  ( $j = 1, 2, \dots, v$ ) is complete, Bayes's formula gives us

$$p_N(H_j) = \frac{p_0(H_j) L_j(N)}{\sum_{k=1}^v p_0(H_k) L_k(N)}, \quad (6.4.4)$$

where

$$L_j(N) = \int p_1[y_1 | \eta_j(x_1, \theta_j)] \cdots p_N[y_N | \eta_j(x_N, \theta_j)]$$

$$p_j[\eta_j(x_N, \theta_j)] d\theta_j \quad (j = 1, 2, \dots, v)$$

From (6.4.3) and (6.4.4) it follows that the increment of information depends not only on the value of the coordinates of the points where the observations were taken but also on the results of the observations.

Therefore, in the future it will make sense to consider the expected value of  $\Delta I$  with respect to the results of the observations. This quantity will be denoted by  $\Delta J(\mathcal{E})$ . A design which maximizes the mean increment of information  $\Delta J(\mathcal{E})$  will be called optimal.

We consider the designs concentrated at one point  $x$ . We assume that the results of observations are independent and normally distributed random variables and that after  $N$  observations the probability of the  $j$ th hypothesis is  $p_N(H_j)$  ( $j = 1, 2, \dots, v$ ). By definition

$$\Delta J(x, N) = \sum_{j=1}^v p_N(H_j) \int p_j(y | x, N) \left\{ \sum_{k=1}^v p_{N+1}(H_k) \ln p_{N+1}(H_k) - \sum_{k=1}^v p_N(H_k) \ln p_N(H_k) \right\} dy, \quad (6.4.5)$$

where

$$p_{N+1}(H_j) = \frac{p_N(H_j) p_j(y | x, N)}{\sum_{l=1}^N p_N(H_l) p_l(y | x, N)}$$

and [cf. (6.2.28)]

$$p_l(y | x, N) = [2\pi s_l(x, N)]^{-1/2} \exp\left(-\frac{1}{2} \frac{[y - \hat{\eta}_l(x, N)]^2}{s_l(x, N)}\right).$$

After a simple transformation of (6.4.5) it is possible to pass to the form

$$\Delta J(x, N) = \sum_{j=1}^v p_N(H_j) \int p_j(y | x, N) \ln \frac{p_j(y | x, N)}{\sum_{l=1}^v p_N(H_l) p_l(y | x, N)}. \quad (6.4.6)$$

In those cases when  $\Delta N$  observations are taken at the point  $x$ , formulas (6.4.5)–(6.4.8) preserve their form if by  $y$  is understood the arithmetical mean  $(\Delta N)^{-1} \sum_{j=1}^{\Delta N} y_j$  of the results of observations at the given point and we set

$$s(x, N) = \lambda^{-1}(x)(\Delta N)^{-1} + d(x, N).$$

The accurate computation of the mean of the increment of information  $\Delta J(x, N)$  presents a fairly complicated problem, since the

improper integrals entering in the second terms of the right-hand side of (6.4.6) are not represented in closed form.

Two paths are possible. Either perform a numerical integration of the necessary expressions, or replace the precise formula (6.4.6) by some reasonable approximation suitable for the given experimental situation.

Numerical integration of the improper integrals is a very cumbersome problem and is hardly useful. It is very often possible for the experimenter, as will be seen from the subsequent exposition, to construct an approximation which is satisfactory in the majority of practical cases [55].

**Theorem 6.4.1.** *The mean increment of information for the design concentrated at a single point of the factor space  $x$  equals*

$$\Delta J(x, N) = -\frac{1}{2} D[\hat{\eta}(x, N)]w + O[w^{3/2}] \quad (6.4.7)$$

where  $w = \lambda(x) \Delta N$ , and

$$\begin{aligned} D[\hat{\eta}(x, N)] = & \sum_{j,k=1}^n p_N(H_j) p_N(H_k) p_N(H_i) [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)] \\ & \times [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)] \end{aligned} \quad (6.4.8)$$

*Proof.* We will consider the product  $\lambda(x) \Delta N$  as some continuous variable  $w$ . It is easy to verify that the function

$$\ln \Sigma(w) = \ln \sum_{k=1}^n p_N(H_k) \frac{p_k(y | x, N)}{p_j(y | x, N)}$$

is continuous for  $w = 0$  and all of its derivatives are finite in any bounded region  $0 \leq w \leq \bar{w}$ .

We expand  $\ln \Sigma(w)$  in a Taylor series at the point  $w = 0$  and truncate it after the first three terms of the series:

$$\ln \Sigma(w) = \ln \Sigma(0) + \frac{\dot{\Sigma}(0)}{\Sigma(0)} w + \frac{\dot{\Sigma}(0) \Sigma(0) - \ddot{\Sigma}(0)}{2 \Sigma^2(0)} w^2 + R(y, N) \quad (6.4.9)$$

where the dot indicates differentiation with respect to  $w$ . Our aim is

to compute the integral (6.4.6), relying on the representation (6.4.9), and to estimate the contribution of the remainder term

$$R(N) = \sum_{j=1}^v p_N(H_j) \int p_j(y | x, N) R_j(y, N) dy. \quad (6.4.10)$$

Differentiating  $\ln \Sigma(w)$ , we obtain for  $w = 0$

$$\Sigma(0) = 1,$$

$$\dot{\Sigma}(0) = -\frac{1}{2} \sum_{k=1}^v p_N(H_k) [2y - \hat{\eta}_k(x, N) - \hat{\eta}_j(x, N)] [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)],$$

$$\ddot{\Sigma}(0) = \frac{1}{4} \sum_{k=1}^v p_N(H_k) [2y - \hat{\eta}_k(x, N) - \hat{\eta}_j(x, N)]^2 [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)]^2.$$

Carrying out a noncomplicated but fairly long calculation it is not difficult to show that

$$\begin{aligned} & \sum_{j=1}^v p_N(H_j) \int p_j(y | x, N) \left[ \ln \Sigma(0) + \frac{\dot{\Sigma}(0)}{\Sigma(0)} w + \frac{\ddot{\Sigma}(0) \dot{\Sigma}(0) - \dot{\Sigma}^2(0)}{2 \Sigma^2(0)} \right] dy \\ &= -\frac{1}{2} \sum_{j,k,l=1}^v p_N(H_j) p_N(H_k) p_N(H_l) \\ & \quad \times [\hat{\eta}_j(x, N) - \hat{\eta}_k(x, N)][\hat{\eta}_j(x, N) - \hat{\eta}_l(x, N)]. \end{aligned} \quad (6.4.11)$$

Representing  $R(y, N)$  in the form

$$R(y, N) = [w^{(n+1)/(n+1)!}][d^{(n+1)}/dw^{(n+1)}] \ln \Sigma(\xi w), \quad 0 \leq \xi \leq 1,$$

and directly integrating it, we may show that for  $w \rightarrow 0$ ,

$$\int R(y, N) p_j(y | x, N) dy \simeq O(w^{3/2}). \quad (6.4.12)$$

In obtaining (6.4.11) and (6.4.12) we used the fact that

$$\int p_j(y | x, N) [y - \hat{\eta}_j(x, N)]^q dy = \begin{cases} 0 & \text{if } q \text{ is odd,} \\ (q-1)! s^{q/2}(x, N) & \text{if } q \text{ is even.} \end{cases} \quad (6.4.13)$$

Equation (6.4.7) follows directly from (6.4.6), (6.4.11), and (6.4.12). The theorem is proved.

More detailed investigation [by comparison with (6.4.12)] of the remainder term shows that

$$\frac{1}{2} D[\hat{\eta}(x, N)] w \gg R(N)$$

when the conditions

$$w^{-1} \gg d_j(x, N) \quad (j = 1, 2, \dots, v) \quad (6.4.14)$$

and

$$w^{-1} \gg D[\hat{\eta}(x, N)] \quad (6.4.15)$$

are satisfied.

Inequality (6.4.14) usually holds for sufficiently large  $N$ , since (compare with Section 6.2)

$$\max_x d_j(x, N) \simeq O\left\{\frac{m_j}{N\lambda(x)}\right\}$$

It is necessary to verify that inequality (6.4.15) is satisfied for each set of functions  $f_j(x)$ .

The quantity  $D[\hat{\eta}(x, N)]$  has a physical sense. Indeed, as is easy to verify, the expression (6.4.8) can be rewritten in the form

$$D[\hat{\eta}(x, N)] = \sum_{j=1}^v p_N(H_j) (\hat{\eta}_j(x, N) - E[\hat{\eta}_j(x, N)])^2 \quad (6.4.16)$$

where

$$E[\hat{\eta}(x, N)] = \sum_{j=1}^v p_N(H_j) \hat{\eta}_j(x, N)$$

If the quantity  $E[\hat{\eta}(x, N)]$  is treated as the expected value of an estimate of the response surface  $\hat{\eta}(x, N)$  at the point  $x$ , then  $D[\hat{\eta}(x, N)]$  can be treated as the dispersion of this surface at the given point.

In this manner the expected increment of information, when (6.4.14) and (6.4.15) are satisfied, is proportional to the scatter (dispersion) of the response surface.

It is curious to note that for sequential estimation of unknown parameters, the smallest determinant of the dispersion matrix of the parameter estimates is proportional to the dispersion  $d(x, N)$ . In Chapter 7, it is shown that this situation is not accidental and has deep significance.

Relying on Theorem 6.4.1 it is possible to carry out a sequential design of experiments according to procedures analogous to those

presented in Sections 6.2 and 6.3. For this, each succeeding observation must be taken where

$$\max_x D[\hat{\eta}(x, N)] \quad (6.4.17)$$

is attained.

EXAMPLE [54]. We consider the following four concurring models:

$$\begin{aligned}\eta_1(x, \theta) &= \exp[-x_1 \exp(\theta_1 - \theta_2 x_2)], \\ \eta_2(x, \theta) &= [1 + x_1 \exp(\theta_1 - \theta_2 x_2)]^{-1}, \\ \eta_3(x, \theta) &= [1 + 2x_1 \exp(\theta_1 - \theta_2 x_2)]^{-1/2}, \\ \eta_4(x, \theta) &= [1 + 3x_1 \exp(\theta_1 - \theta_2 x_2)]^{-1/3}.\end{aligned}$$

The form of the model and the numerical data corresponding to them are adopted from [55]. The domain of possible observations is determined by the inequalities  $0 \leq x_1 \leq 150$ ;  $x_2 = [(1/T) - (1/525)]$ ,  $450 \leq T \leq 600$ .

For these models a sequential design defined by (6.4.17) was carried out in determining a suitable model. For the true model the second was chosen with  $\theta_1 = -3.53$  and  $\theta_2 = 5000$ . The error of measurement was assumed to be equal to  $b = 0.05$ .

Since the parametrization is nonlinear, the estimate  $\hat{\eta}_j(x, \theta)$  and its dispersion  $d_j(x, N)$  is sought after each observation by methods presented in Section 1.4, and their value is directly substituted in (6.4.8). Note that after only six observations the following results say with great certainty that the second model is the true one (for details see Table 9).

**Table 9**  
Results of Sequential Design Using Eq. (6.4.17)

$N$	$x_1$	$T$	$y$	$p_{N(H_1)}$	$p_{N(H_2)}$	$p_{N(H_3)}$	$p_{N(H_4)}$
1	25	575	0.396				
2	25	475	0.723				
3	125	475	0.422				
4	125	575	0.1297	0.001	0.390	0.612	0.002
5	4.24	600	0.705	0.26( $10^{-3}$ )	0.51	0.487	0.28( $10^{-2}$ )
6	67.6	600	0.143	1.0( $10^{-5}$ )	0.875	0.125	0.1( $10^{-4}$ )
7	76.4	600	0.073	0.6( $10^{-5}$ )	0.997	0.003	0.15( $10^{-8}$ )
8	76.3	600	0.118	0.4( $10^{-6}$ )	0.9996	0.4( $10^{-3}$ )	$\sim 10^{-11}$
9	9.24	573	0.595	$\sim 10^{-10}$	0.9999	0.78( $10^{-4}$ )	$\sim 10^{-13}$
10	150	450	0.487	$\sim 10^{-10}$	0.9999	0.45( $10^{-4}$ )	$\sim 10^{-14}$

priming experiment

# 7

## Generalized Criteria of Optimality

### 7.1. Experiments Minimizing Generalized Loss

I. The methods considered in the preceding chapters may be applied in cases when either unknown parameters are being estimated or when a test of hypotheses is being conducted

The corresponding experimental designs permit us to effectively extract information, either only about the unknown parameters, or only about the hypotheses being investigated. The ordering of the various designs, obtained from Diagram I (cf. Introduction) is conducted by the experimenter using semi-intuitive considerations and relying basically on the theoretical position of the given branch of science and on the results of analogous investigations.

If the analytic form of the investigated model or dependence is not very complicated and a visual interpretation of the results is possible, then such methods usually lead to good results.

When the concurring models have a complicated analytic form or only a theoretical interpretation of the results is possible, the experimenter may find himself in a position where his intuition is rather weak.

In this case it is natural to turn to criteria of optimality of experiments which would permit one to combine the problem of finding

the true model and the problem of determining the estimates of the parameters.

There are two possible paths for joining these criteria. The first considers criteria depending on a measure, which is a composition of using earlier measures of the loss for discriminating experiments (cf. Chapter 6), and measures of the loss for determining or making precise the estimates of the unknown parameters (cf. Chapter 1).

The second turns toward a new measure of accuracy (which in its own right uniquely determines the loss) of the results of an experiment, simultaneously taking into account information used in discriminating hypotheses and in determining estimates of sought parameters.

II. We turn to the first type of generalized criteria. The second will be considered in the following section.

Let the loss, in the case of an incorrect decision, be characterized by the quantity

$$\mathcal{R} = W_1 \mathcal{D} + W_2 \mathcal{L} \quad (7.1.1)$$

where  $W_1$  and  $W_2$  are some weight multipliers,  $\mathcal{D}$  is the loss for an incorrect decision in discriminating hypotheses, and  $\mathcal{L}$  is the loss for not determining the estimates of the unknown parameters with sufficient accuracy.

The representation of the loss  $\mathcal{R}$  in the form (7.1.1) is a compromise between two, very frequently, contradictory requirements. Indeed,  $\mathcal{R}$  as a function of  $\mathcal{D}$  and  $\mathcal{L}$  must have a simple analytic form and must sufficiently well describe the real situation.

As measures  $\mathcal{D}$  and  $\mathcal{L}$ , depending on the needs of the experimenter, can be taken to be any measure of accuracy, considered in the preceding chapters.

In the process of observation the weight multipliers, generally speaking, must be variable. Indeed, if the obtained results say that the probability of one of the models is close to one, then the relationship of the weight multipliers must be such that

$$W_1 \mathcal{D} \ll W_2 \mathcal{L}. \quad (7.1.2)$$

If from the obtained results it follows that the probability of all (or part) of the models are close to one another, then the reverse inequality must hold

$$W_1 \mathcal{D} \gg W_2 \mathcal{L}. \quad (7.1.3)$$

As one of the possible realizations of (7.1.1) one may, for example, choose a function defined by means of the multipliers [55]

$$W_1 = [\lambda(1 - p_N)(v - 1)]^{\frac{1}{\lambda}}, \quad W_2 = 1 - W_1, \quad (7.1.4)$$

where  $p_N = \max_j p_N(H_j)$ ,  $p_N(H_j)$  is the probability of the  $j$ th hypothesis after  $N$  observations ( $j = 1, 2, \dots, v$ ) and  $\lambda$  is a constant between 0 and  $\infty$ .

The multipliers (7.1.4) observe the properties indicated above. When none of the hypotheses is preferred, that is,  $p_N(H_j) = v^{-1}$ ,  $W_1$  equals unity and  $W_2$  equals zero, and the problem reduces to the problem of discriminating hypotheses. On the other hand, when one of the hypotheses is known to be true [ $p_N(H_j) = 1$ ], the problem reduces to the problem of seeking an estimate of parameters.

For the intermediate cases,  $W_1$  is monotonically decreasing in  $p_N$ , as is shown in Fig. 33. The speed of decrease is determined by the value  $\lambda$ , which is chosen by the experimenter. For any fixed value of  $p_N$  the weight function  $W_1$  decreases with an increase in  $\lambda$ . Therefore

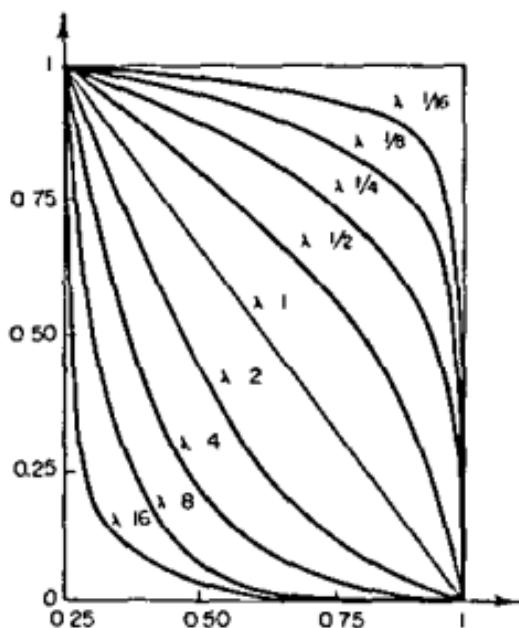


Fig. 33 Behavior of the weight function (7.1.4) for various  $\lambda$  [55]

the quantity  $\lambda$  can decrease the role of the problem of discriminating hypotheses and increase the role of the problem of finding estimates of parameters.

We consider a sequential procedure for designing experiments analogous in its nature to the sequential procedure investigated in Chapters 4 and 6. At each of the  $N$  stages a point  $x_N$  is sought corresponding to the maximum of the

$$\max_x [\mathcal{R}(N) - \mathcal{R}(N+1)]. \quad (7.1.5)$$

The last observation (or group of observations) must be allocated to the point  $x_N$ . A data reduction is then conducted and all operations are repeated.

The choice of the weight multipliers (7.1.4) stipulates that the losses  $\mathcal{D}$  and  $\mathcal{L}$  be expressed in the same units (for example, monetary).

## 7.2. The Information Approach to the General Problem of Seeking the True Mathematical Model

I. The methods developed above for designing experiments rely critically on the possibility of describing the results through estimates of unknown parameters  $\hat{\theta}$  and their covariance matrix  $D(\hat{\theta})$ .

The advantages of such a description of the results of an experiment were presented in Chapter 1. Unfortunately, for nonlinear parametrization, the representation of the results in terms of  $\hat{\theta}$  and  $D(\hat{\theta})$  is not always possible. Usually, this happens when the number of observations is not large [the sample size  $Y_n$  is small and inequality (1.4.9) is not satisfied] or when the sum of the weighted squares of the deviations  $\sum_{i=1}^n w_i(y_i - \eta(x_i, \theta))^2$  has several, nearly equal local minima [cf. the remarks to (1.4.3)]. In these cases, for a description of the results, it is natural to turn to the *a posteriori* probability distribution function  $p(\theta)$  in the space of parameters. For its construction it is necessary to know the analytic form of the conditional density function  $p[y | \eta(x, \theta)]$  of the results of the observations. Recall that for the construction of the best linear estimates it is sufficient to know only the first and second moments of the function  $p[y | \eta(x; \theta)]$ . Therefore, the description of the results of an experiment, using  $p(\theta)$ , requires more prior information than the description of these results using the techniques of best linear estimates.

Since  $p(\theta)$ , in the general case, is defined on a multidimensional surface, the comparison of the results of various experiments directly from the form of  $p(\theta)$  is very cumbersome from the conceptual point of view. Even in those cases when the function  $p(\theta)$  is defined on a space of small dimension, it is necessary to deal with situations where it is difficult to state a preference for one or another experiment, relying directly on the form of  $p(\theta)$ . One of these cases is presented in Fig. 34. Earlier, in comparing the results of experiments, each

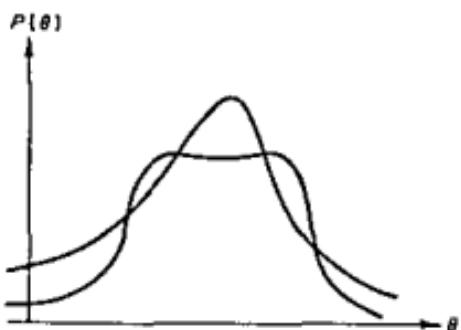


Fig. 34 Posterior density function for two distinct experiments

matrix  $D(\theta)$  was assigned some scalar quantity. Now each surface  $p(\theta)$  must be assigned a scalar quantity. It is evident that the indicated quantity must satisfy a series of natural requirements, for example, it must increase with the amount of experimental data.

For such a quantity one may choose the entropy measure of information, which has already been mentioned in connection with the design of experiments for discriminating hypotheses.

If the density of the distribution in the space of parameters is  $p(\theta)$ , then the entropy (measure of lack of order) is equal to

$$I[p(\theta)] = - \int p(\theta) \ln p(\theta) d\theta \quad (7.2.1)$$

Let an experiment  $\mathcal{E}$  be conducted with observations  $y' = \{y_1, y_2, \dots, y_N\}$ . Then the measure of acquired knowledge will be the quantity

$$\Delta I[\mathcal{E}, p_0(\theta)] = - \int p_0(\theta) \ln p_0(\theta) d\theta + \int p(\theta | y) \ln p(\theta | y) d\theta, \quad (7.2.2)$$

where  $p(\theta | y)$  is the *a posteriori* density of the parameters

In those cases when there are several hypotheses and we are interested in choosing the best of them, as in estimating unknown parameters, (7.2.2) takes on the form

$$\begin{aligned} \Delta I[\mathcal{E}, p_0(\theta_j)] &= - \sum_{j=1}^v \int p_0(\theta_j) \ln p_0(\theta_j) d\theta_j \\ &\quad + \sum_{j=1}^v \int p(\theta_j | \mathbf{y}) \ln p(\theta_j | \mathbf{y}) d\theta_j, \end{aligned} \quad (7.2.3)$$

where  $j = 1, 2, \dots, v$ . If a space  $\Omega$  is introduced into the considerations, each point of which is defined by means of the index  $j$  and the vector  $\theta_j$ , then (7.2.3) may be rewritten formally in the form

$$\Delta I[\mathcal{E}, p_0(\Theta)] = - \int_{\Omega} p_0(\Theta) \ln p_0(\Theta) d\Theta + \int_{\Omega} p(\Theta | \mathbf{y}) \ln p(\Theta | \mathbf{y}) d\Theta, \quad (7.2.4)$$

where  $\Theta$  is a vector corresponding to a point of the space  $\Omega$ .

The quantity  $\Delta I[\mathcal{E}, p_0(\Theta)]$  is called the measure of the amount of information (or simply information) gained in the experiment  $\mathcal{E}$  with *a priori* density  $p_0(\Theta)$ .

We will say that the results of the experiment  $\mathcal{E}_1$  are better than the results of experiment  $\mathcal{E}_2$  if

$$\Delta I[\mathcal{E}_1, p_0(\Theta)] > \Delta I[\mathcal{E}_2, p_0(\Theta)]. \quad (7.2.5)$$

**II.** We consider the properties of the information defined by (7.2.4). From (7.2.4) it follows that  $\Delta I[\mathcal{E}, p(\Theta)]$  depends not only on the design of the experiment but also on the results of the experiment. Therefore, in planning we will rely on the expectation of this quantity with respect to the results of the observations:

$$\begin{aligned} \mathcal{I}[\mathcal{E}(N), p_0(\Theta)] &= \int p(\mathbf{y}) \Delta I[\mathcal{E}, p_0(\Theta)] d\mathbf{y} \\ &= E_{\mathbf{y}}\{\Delta I[\mathcal{E}, p_0(\Theta)]\}, \end{aligned} \quad (7.2.6)$$

where

$$p(\mathbf{y}) = \int p(\mathbf{y} | \Theta) p_0(\Theta) d\Theta = E_{\Theta}[p(\mathbf{y} | \Theta)] \quad (7.2.7)$$

and  $\mathcal{E}(N)$  is the design of the experiment  $\mathcal{E}$ , consisting of  $N$  observations.

According to Bayes's formula,

$$p(\theta | y) = \frac{p(y | \theta) p_0(\theta)}{p(y)} \quad (7.2.8)$$

From (7.2.6) and (7.2.8) it follows that

$$\mathcal{I}[\mathcal{E}(N) p_0(\theta)] = \int p(y, \theta) \ln \frac{p(y, \theta)}{p(y) p_0(\theta)} dy d\theta \quad (7.2.9)$$

where

$$\mathcal{I}[\mathcal{E}(N) p_0(\theta)] = E_r F_\theta \ln \frac{p(y | \theta)}{p(y) p_0(\theta)}, \quad (7.2.10)$$

and

$$p(y | \theta) = p(y | \theta) p_0(\theta) = p(\theta | y) p(y) \quad (7.2.11)$$

In deducing (7.2.9) we used the fact that

$$\int p(y) p(\theta | y) dy = p_0(\theta)$$

Relying on (7.2.9) it is not difficult to verify that the expected increment of information  $\mathcal{I}[\mathcal{E}(N) p(\theta)]$  is invariant with respect to any arbitrary single-valued transformation in the space  $\theta$

**III** The future results are based on Bayes's theorem and the properties of the logarithm function (in particular, on its concavity). In those places where no ambiguity may arise the term  $p_0(\theta)$  in the notation of the expected information will be omitted  $\mathcal{I}[\mathcal{E}(N)] = \mathcal{I}[\mathcal{E}(N) p_0(\theta)]$

We consider the basic properties of the entropy measure of information (cf. basically [56])

**Theorem 7.2.1** *The expected increment of information for any design  $\mathcal{E}(N)$  is nonnegative*

$$\mathcal{I}[\mathcal{E}(N)] \geq 0 \quad (7.2.12)$$

*the equality sign holds only if  $p(y | \theta)$  does not depend on  $\theta$*

*Proof.* It is known that [57]

$$\begin{aligned} & \int f(y, \Theta) p_0(\Theta) p(y) \ln f(y, \Theta) dy d\Theta \\ & \geq \int f(y, \Theta) p_0(\Theta) p(y) dy d\Theta \ln \left\{ \frac{\int f(y, \Theta) p(y) p_0(\Theta) dy d\Theta}{\int p(y) p_0(\Theta) dy d\Theta} \right\}, \end{aligned} \quad (7.2.13)$$

and equality is satisfied only if  $f(y, \Theta) = \text{const}$ , excluding perhaps a set of measure zero.

Setting

$$f(y, \Theta) = \frac{p(y, \Theta)}{p(y) p_0(\Theta)},$$

we rewrite the expression for the expected increment of information in the form

$$\mathcal{I}[\mathcal{E}(N)] = \int f(y, \Theta) p_0(\Theta) p(y) \ln f(y, \Theta) dy d\Theta. \quad (7.2.14)$$

From (7.2.13) and (7.2.14)

$$\mathcal{I}[\mathcal{E}(N)] \geq \int f(y, \Theta) p(y) p_0(\Theta) dy d\Theta \ln \frac{\int f(y, \Theta) p(y) p(\Theta) dy d\Theta}{\int p(y) p(\Theta) dy d\Theta}.$$

Since

$$\int f(y, \Theta) p(y) p_0(\Theta) dy d\Theta = 1$$

and

$$\int p(y) p_0(\Theta) dy d\Theta = 1,$$

then the logarithm on the right-hand side of the inequality equals zero and

$$\mathcal{I}[\mathcal{E}(N)] \geq 0,$$

which is what we were required to prove.

Theorem 7.2.1 says that any experiment is informative on the average, that is, helps the experimenter clarify the state of the system being studied. We note however, that the increment of the information  $\Delta I[\mathcal{E}, p_0(\Theta)]$  is not required to be positive.

In those cases when the results of observations take on values with

small probability (an accidental outcome) the true  $\Theta$  can be ascribed a smaller probability than before [cf (7.2.8)] and the entropy of the system increases.

We assume that the experiment  $\mathcal{E}$  consists of two sequentially conducted experiments  $\mathcal{E}_1$  and  $\mathcal{E}_2$ . Let the results of the experiment  $\mathcal{E}_1$  be  $y_1$  and the results of experiment  $\mathcal{E}_2$  be  $y_2$ .

We consider the quantity  $I[\mathcal{E}(N_2) | p(\Theta | y_1)]$ . Since the density of the probability distribution  $p(\Theta | y_1)$  depends on  $y_1$ , then the expected increase of information for the design  $\mathcal{E}(N_2)$  is obtained after the experiment  $\mathcal{E}_1$  and depends on the results of this experiment.

The average of this quantity with respect to  $y_1$  gives the mean increment of information obtained in the experiment carried out according to the design  $\mathcal{E}(N_2)$  which is preceded by the experiment, conducted according to the design  $\mathcal{E}(N_1)$ . We denote

$$E_{y_1}\{I[\mathcal{E}(N_2) | p(\Theta | y_1)]\} \quad \text{by} \quad I[\mathcal{E}(N_2) | \mathcal{E}(N_1)]$$

Repeating the proof of Theorem 7.2.1 practically word for word, it is not difficult to obtain that

$$I[\mathcal{E}(N_2) | \mathcal{E}(N_1)] \geq 0$$

in which case equality holds if and only if  $p(y_2 | \Theta, y_1)$  does not depend on  $\Theta$ , except possibly on a set of measure zero.

### Theorem 7.2.2

$$I[\mathcal{E}(N_1)] + I[\mathcal{E}(N_2) | \mathcal{E}(N_1)] = I[\mathcal{E}(N_1 + N_2)] \quad (7.2.15)$$

where  $\mathcal{E}(N_1 + N_2)$  is the design of the experiment  $\mathcal{E}$ .

*Proof* In agreement with (7.2.10)

$$I[\mathcal{E}(N_1)] = E_{y_1} E_\Theta \ln \frac{p(y_1 | \Theta)}{p(y_1) p_0(\Theta)} = E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_1 | \Theta)}{p(y_1) p_0(\Theta)} \quad (7.2.16)$$

By definition

$$\begin{aligned} I[\mathcal{E}(N_2) | \mathcal{E}(N_1)] &= E_{y_1}\{I[\mathcal{E}(N_2) | p(\Theta | y_1)]\} \\ &= E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_2 | \Theta | y_1)}{p(y_2 | y_1) p(\Theta | y_1)} \end{aligned} \quad (7.2.17)$$

Combining (7.2.16) and (7.2.17) and using (7.2.7) and (7.2.8) we obtain

$$\begin{aligned} E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_2, \Theta | y_1)}{p(y_2 | y_1) p(\Theta | y_1)} + E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_1, \Theta)}{p(y_1) p_0(\Theta)} \\ = E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_1, y_2, \Theta)}{p(y_1, y_2) p_0(\Theta)}. \end{aligned}$$

By definition, the right-hand side of the given equality is the expected increment of information in the experiment  $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$ , which proves the theorem.

Theorem 7.2.2 says that the expected information obtained in the experiment is an additive quantity.

Shannon (cf. [58]; the proof was conducted for the discrete case) showed that when Eq. (7.2.16) is used as a definition, then the function

$$I = - \int p(\Theta) \ln p(\Theta) d\Theta + c,$$

where  $c$  is some constant, is the unique function up to constant multipliers satisfying (7.2.15) and the condition of continuity.

We assume that the experiments  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are independent, that is,

$$p(y_1, y_2 | \Theta) = p(y_1 | \Theta) p(y_2 | \Theta) \quad (7.2.18)$$

for any  $\Theta$ . The following assertion holds.

**Theorem 7.2.3.** *For the experiments  $\mathcal{E}_1$  and  $\mathcal{E}_2$*

$$\mathcal{I}[\mathcal{E}(N_2) | \mathcal{E}(N_1)] \leq \mathcal{I}[\mathcal{E}(N_2)]; \quad (7.2.19)$$

*equality holds if and only if  $y_1$  and  $y_2$  are independent, that is,*

$$p(y_1, y_2) = p(y_1) p(y_2). \quad (7.2.20)$$

*Proof.* From (7.2.16), (7.2.17), and the independence condition we have

$$\begin{aligned} & \mathcal{I}[\mathcal{E}(N_2)] - \mathcal{I}[\mathcal{E}(N_2) | \mathcal{E}(N_1)] \\ &= E_{y_2} E_\Theta \ln \frac{p(y_2 | \Theta)}{p(y_2)} - E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_2 | \Theta)}{p(y_2 | y_1)} \\ &= E_{y_1} E_{y_2} E_\Theta \ln \frac{p(y_2 | y_1)}{p(y_2)} \\ &= E_{y_1} E_{y_2} \ln \frac{p(y_2 | y_1)}{p(y_1) p(y_2)} \end{aligned}$$

The last expression will be identical to (7.2.10) if  $y_2$  and  $y_1$  are replaced by  $y$  and  $\theta$ , respectively. By Theorem 7.2.1 it is nonnegative, so that

$$\mathcal{I}[\mathcal{E}(N_t)] - \mathcal{I}[\mathcal{E}(N_2) | \mathcal{E}(N_1)] \geq 0$$

The equality sign holds only if  $p(y_2 | y_1) = p(y_2)$ . The theorem is proved.

**Corollary 1.** If the experiments  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are independent, then

$$\mathcal{I}[\mathcal{E}(N_1)] + \mathcal{I}[\mathcal{E}(N_t)] \geq \mathcal{I}[\mathcal{E}(N_1 + N_t)] \quad (7.2.21)$$

Equality holds only if  $y_1$  and  $y_2$  are independent [cf. (7.2.20)].

Inequality (7.2.21) follows from Theorems 7.2.2 and 7.2.3.

$$\mathcal{I}[\mathcal{E}(N_1)] + \mathcal{I}[\mathcal{E}(N_t)] \geq \mathcal{I}[\mathcal{E}(N_1)] + \mathcal{I}[\mathcal{E}(N_t) | \mathcal{E}(N_1)] = \mathcal{I}[\mathcal{E}(N_1 + N_t)]$$

Theorem 7.2.3 and its corollary say, in particular, that in conducting experiment  $\mathcal{E}(N_2)$ , according to the design  $\mathcal{E}(N_2) = \mathcal{E}(N_1)$ , after the experiment  $\mathcal{E}(N_1)$ , we will obtain an increment of information always less than that from the experiment  $\mathcal{E}(N_1)$ .

This property is in good agreement with the intuitive objectives of the experimenter in allocating observations at various points of the factor space or in conducting principally distinct experiments in order to satisfy (7.2.20). (It is possible with some degree of approximation.) Then (7.2.21) becomes an equality.

Relying on (7.2.23) it is possible to easily verify that for identical designs  $\mathcal{E}(N_1) = \mathcal{E}(N_2) = \dots = \mathcal{E}(N_k)$  the expected information is a convex increasing function of  $N$ .

**IV.** Having investigated the general properties of information, we now consider the design of experiments. Below we will assume that the results of the experiments are distributed according to the normal law

$$p(y | \theta) = p[y | \eta_i(x, \theta_i)] = [2\pi b^2(x)]^{-1/2} \exp\left(-\frac{1}{2} \frac{[y - \eta_i(x, \theta_i)]^2}{b^2(x)}\right) \quad (7.2.22)$$

where, as usual, the observations are assumed to be independent

$$p[y_1, y_2 | \eta_j(x, \theta_j)] = p[y_1 | \eta_j(x, \theta_j)] p[y_2 | \eta_j(x, \theta_j)].$$

If at a point  $x$ ,  $\Delta N$  observations were taken, then in (7.2.22) we replace  $y$  by  $y = (\Delta N)^{-1} \sum_{j=1}^{\Delta N} y_j$  and  $b(x)$  by  $[\lambda(x) \Delta N]^{-1/2}$ . Our aim is to construct designs  $\mathcal{E}(N)$  maximizing the mean increment of information:

$$\mathcal{I}[\mathcal{E}(N), p(\Theta)] = \max_{\mathcal{E}(N)} \mathcal{I}[\mathcal{E}(N), p(\Theta)].$$

We consider one-point designs. That is, after  $N$  observations let an experiment be conducted according to a design whose spectrum consists of a single point  $x$ . Then

$$\begin{aligned} p(\Theta, N + \Delta N) &= \frac{p[y | \eta_j(x, \theta_j)] p(\theta_j, N)}{\sum_{j=1}^v \int p[y | \eta_j(x, \theta_j)] p(\theta_j, N) d\theta_j} \\ &= \frac{p[y | \eta_j(x, \theta_j)] p(\theta_j, N)}{p(y | x, N)}. \end{aligned} \quad (7.2.23)$$

In (7.2.23) it was assumed that at the point  $x$ ,  $\Delta N$  observations are conducted.

From (7.2.9) and (7.2.23)

$$\mathcal{I}[\mathcal{E}(\Delta N), p(\Theta, N)] = \sum_{j=1}^v p[y | \eta_j(x, \theta_j)] p(\theta_j, N) \ln \frac{p[y | \eta_j(x, \theta_j)]}{p(y | x, N)} dy d\theta_j. \quad (7.2.24)$$

Since the conditional density function of the results of the observations depends only on  $\eta_j(x, \theta_j)$ , then

$$\begin{aligned} \mathcal{I}[\mathcal{E}(\Delta N), p(\Theta, N)] &= \mathcal{I}\{\mathcal{E}(\Delta N), p[\eta(x), N]\} \\ &= \int p[y | \eta(x)] p[\eta(x), N] \ln \frac{p[y | \eta(x)]}{p(y | x, N)} dy d\eta(x), \end{aligned} \quad (7.2.25)$$

where  $\eta(x)$  is a possible value of the response surface at the point  $x$ . In what follows, where a second interpretation is not possible, the argument  $x$  will be omitted.

The formula (7.2.25) can be obtained by means of a direct computation, indeed,

$$\begin{aligned} & \int p(y|\eta) p(\eta, N) \ln \frac{p(y|\eta)}{p(y, N)} dy d\eta \\ &= \int p(y|\eta) \left\{ \sum_{j=1}^s \int p(\theta_j, N) \right. \\ &\quad \times \delta[\eta(x) - \eta_j(x, \theta_j)] d\theta_j \left. \right\} \ln \frac{p(y|\eta)}{p(y, N)} dy d\eta \\ &= \sum_{j=1}^s \int p[y + \eta(x, \theta_j)] p(\theta_j, N) \ln \frac{p[y + \eta(x, \theta_j)]}{p(y, N)} dy d\theta_j \\ &= \mathcal{I}[\mathcal{E}(\Delta N) \ p(\theta, N)] \end{aligned}$$

The transformations carried out above rely on the well known formula (cf., for example, [5])

$$p(\eta) = \int p(\theta) \delta[\eta - \eta(\theta)] d\theta \quad (7.2.26)$$

Relying on (7.2.26) it is not difficult also to verify that

$$p(y|N) = \int p(y|\eta) p(\eta|N) d\eta \quad (7.2.27)$$

**Theorem 7.2.4** *The expected increment of information for the design concentrated at one point  $x$  of the factor space is equal to*

$$\mathcal{I}[x \ p(\theta_j, N)] = \frac{1}{2} D[\eta(x)|N]w + O(\xi w^{3/2}) \quad (7.2.28)$$

where  $w = \lambda(x) \Delta N$ ,  $0 \leq \xi \leq 1$ , and

$$D[\eta(x)|N] = E[\eta(x) - E\eta(x)]^2 \quad (7.2.29)$$

*Proof* From (7.2.22) and (7.2.25)

$$\begin{aligned} & \mathcal{I}[\mathcal{E}(\Delta N) \ p(\theta_j, N)] \\ &= \int p(\eta|N) p(y|\eta) \ln \left[ \frac{\exp[(2y\eta - \eta)^2 w/2]}{\int \exp[(2y\eta - \eta)^2 w/2] p(\eta) d\eta} \right] dy d\eta \end{aligned} \quad (7.2.30)$$

Expanding the logarithms entering in (7.2.30) in a series in  $w = \lambda(x) \Delta N$  and keeping terms up to the second derivative inclusive, we may obtain (7.2.28) in a manner completely analogous to Theorem 6.3.1.

The quantity  $D[\eta(x), N]$  can be treated as the dispersion of the random variable  $\eta(x)$ .

We remark that in Eq. (7.2.28) the concrete form  $p(\eta, N)$  is not used. Therefore the results of Theorem 7.2.4 apply with arbitrary parametrization of the response surface  $\eta_j(x, \theta_j)$ . In this case the fundamental difficulty in computing  $D[\eta(x), N]$  consists in computing the integral (7.2.26). If the integral (7.2.26) is known, then the further computations do not present any difficulties.

In what follows, two cases will be considered. The majority of practical problems may be reduced to these cases and the integral (7.2.26) can be computed in a closed form.

V. Let there be  $j$  hypotheses, where

$$\eta_j(x, \theta_j) = \sum_{\alpha=1}^{m_j} \theta_{j\alpha} f_{j\alpha}(x), \quad (7.2.31)$$

and let  $N$  observations be taken. Then, under the assumption that the *a priori* probabilities of the values  $\theta_j$  ( $j = 1, 2, \dots, v$ ) are equal, we have

$$\begin{aligned} p(\theta_j, N) &= \frac{\prod_{i=1}^N p[y_i | \eta_j(x, \theta_j)] p(\theta_j)}{\sum_{j=1}^v \int \prod p[y_i | \eta_j(x, \theta_j)] p(\theta_j) d\theta_j} \\ &\sim \exp[-\frac{1}{2} S_j(N) - \frac{1}{2}(\hat{\theta}_j - \theta_j)' D_j^{-1}(N)(\hat{\theta}_j - \theta_j)]. \end{aligned} \quad (7.2.32)$$

In (7.2.32) the representation (6.2.29) was used. Recall that

$$S_j(N) = \sum_{i=1}^n w_i [y_i - \eta_j(x_i, \hat{\theta}_j)]^2,$$

where  $\hat{\theta}_j$  is the best linear estimate, constructed after  $N$  observations, under the assumption of the truth of the  $j$ th hypothesis, and  $D_j(N)$  is their dispersion matrix.

Using the normalization condition

$$\sum_{j=1}^v \int p(\theta_j, N) d\theta_j = 1,$$

we rewrite (7.2.32) in the form

$$p(\theta_j, N) = p_N(H_j) p_j(\theta_j, N), \quad (7.2.33)$$

where

$$p_N(H_j) = \frac{(2\pi)^{-m_j/2} |D_j(N)|^{-1/2} \exp[-\frac{1}{2} S_j(N)]}{\sum_{k=1}^n (2\pi)^{-m_k/2} |D_k(N)|^{-1/2} \exp[-\frac{1}{2} S_k(N)]},$$

and

$$p_j(\theta_j, N) = (2\pi)^{-m_j/2} |D_j(N)|^{-1/2} \exp[-\frac{1}{2} (\theta_j - \hat{\theta}_j) D^{-1}(N) (\theta_j - \hat{\theta}_j)]$$

The weight multiplier  $p_N(H_j)$  is not very different from the probability of the  $j$ th hypothesis (cf. Section 6.4). If the hypotheses are not equally likely *a priori*, then in the numerator and in each term of the denominator of the right-hand side (7.2.33) it is necessary to add the multiplier  $p_0(H_j)$ , where  $p_0(H_j)$  is the *a priori* probability of the  $j$ th hypothesis.

Setting (7.2.33) in (7.2.26), we obtain

$$p(\eta) = \sum_{j=1}^n p_N(H_j) [2\pi d_j(x, N)]^{-1/2} \exp\left(-\frac{1}{2} \frac{[\eta - \hat{\eta}_j(x, \theta_j)]^2}{d_j(x, N)}\right) \quad (7.2.34)$$

where  $\hat{\eta}_j(x, \theta_j) = f'(x) \hat{\theta}_j$  and  $d_j(x, N)$  is the dispersion of the estimate  $\hat{\eta}_j(x, \theta_j)$ .

Relying on (7.2.34) we compute the value  $D[\eta(x) | N]$ . From the definition of the mean

$$\begin{aligned} E(\eta) &= \int \eta p(\eta) d\eta \\ &= \sum_{j=1}^n p_N(H_j) [2\pi d_j(x, N)]^{-1/2} \int \eta \exp\left(-\frac{1}{2} \frac{[\eta - \hat{\eta}_j(x, \theta_j)]^2}{d_j(x, N)}\right) d\eta \\ &= \sum_{j=1}^n p_N(H_j) \hat{\eta}_j(x, \theta_j) \end{aligned}$$

Setting the obtained expression in (7.2.29) we obtain

$$\begin{aligned} D[\eta(x) | N] &= \sum_{j=1}^n p_N(H_j) [2\pi d_j(x, N)]^{-1/2} \\ &\quad \times \int \left[ \eta - \sum_{k=1}^n p_N(H_k) \hat{\eta}_k(x, \theta_k) \right]^2 \exp\left(-\frac{1}{2} \frac{[\eta - \hat{\eta}_j(x, \theta_j)]^2}{d_j(x, N)}\right) d\eta \\ &= \sum_{j=1}^n p_N(H_j) d_j(x, N) + \sum_{j=1}^n p_N(H_j) [\hat{\eta}_j(x, \theta_j) - \sum_{k=1}^n p_N(H_k) \hat{\eta}_k(x, \theta_k)]^2 \end{aligned}$$

Conducting a detailed analysis of the remainder term in (7.2.28), by means on a straightforward but cumbersome computation, it is possible to show that this member is small in comparison with  $\frac{1}{2}D[\eta(x), N]w$  for [cf. from (6.4.20) and (6.4.21)]

$$w^{-1} \ll d_j(x, N)$$

and

$$w^{-1} \ll \sum_{j=1}^v p_N(H_j)[\hat{\eta}_j(x, \theta_j) - \sum_{k=1}^v p_N(H_k) \hat{\eta}_k(x, \theta_k)]^2.$$

In this manner, if the inequalities introduced above are valid, then [54]

$$\mathcal{I}[A\mathcal{E}, p(\theta, , N)] \simeq \frac{1}{2} D[\hat{\eta}(x), N]. \quad (7.2.35)$$

From (7.2.35) it follows that the greatest information is obtained for the observation taken where the generalized dispersion has the largest value (cf. Chapter 3).

The larger the generalized dispersion the less we know about each of the response surfaces and the greater the distance between them in the sense of the metric

$$\sum_{j=1}^v p_N(H_j)[\hat{\eta}_j(x, \theta_j) - \sum_{k=1}^v p_N(H_k) \hat{\eta}_k(x, \theta_k)]^2.$$

The results obtained in this section generalize to the case of non-linear parametrization if for each of the hypotheses the conditions of Section 1.4 are satisfied. In this case, in all of the formulas, it is necessary to replace  $f_{j\alpha}(x)$  by  $(\partial/\partial\theta_{j\alpha}) \eta(x, \theta_{j\alpha})|_{\theta_j=\theta}$ .

**VI.** Let the parametrization of the response surface  $\eta(x, \theta)$  be such that at some stage of the experiment the function  $p(\theta, N)$  has (cf. Fig. 35) some local maximum, of approximately the same height [this is equivalent to the fact that Eq. (1.4.4) has several roots]. It is necessary to deal with situations of this type, for example, in the reduction of experimental data in experiments in the scatter of elementary particles [59].

In the general case the integrals (7.2.26) and (7.2.27) cannot be obtained in closed form. Numerical integration of these quantities is an extremely cumbersome problem for large-dimensional  $\theta$ .

One possible realistic approach [60, 61] to the solution of this problem is to replace the function  $p(\theta, N)$ , by the function  $\tilde{p}(\theta, N)$ ,

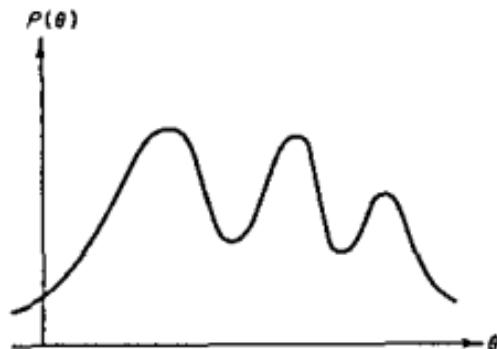


Fig. 35. An example of multi-modal posterior density function

which has a simple analytic form [from the point of view of computing the integrals (7.2.28) and (7.2.29)] and at the same time approximates  $p(\theta, N)$  sufficiently well. The approximation may, for example, be in the sense of the metric  $\int [p(\theta, N) - \tilde{p}(\theta, N)]^2 d\theta$ , which in turn implies that the integral  $I[p(\theta, N)]$  is approximately equal to  $I[\tilde{p}(\theta, N)]$ .

If the extremal points are sufficiently far from one another, then the function  $p(\theta, N)$  can be approximated by the superposition of normal distributions

$$\tilde{p}(\theta, N) = \sum_{j=1}^v p_N(H_j) p_j(\theta, N) \quad (7.2.36)$$

Here

$$p_N(H_j) = \frac{|D_j(N)|^{1/2} p_j(\theta, N)}{\sum_{k=1}^v |D_k(N)|^{1/2} p_k(\theta, N)},$$

the matrix  $D_j(N)$  is defined by formula (1.4.11) with

$$f_{ji}(v) = \left. \frac{\partial \eta(x, \theta)}{\partial \theta_i} \right|_{\theta=\theta_j}$$

Here also,  $\theta_j$  are the coordinates of the  $j$ th local maxima of the function  $p(\theta, N)$ , and  $v$  is the number of these maxima.

We represent (7.2.36) using quasi-linearity in the neighborhood of each of the maxima

$$\eta(x, \theta) \approx \eta(x, \theta_j) + f_{ji}(v)(\theta - \theta_j) \quad (7.2.37)$$

Directly differentiating (7.2.8) under conditions satisfied by (7.2.37) it is possible for us to verify that in the region close to the extremal points  $\theta_j$  (which is the main concern of the experimenter), the func-

tions  $p(\theta, N)$  and  $\sum_{j=1}^v p_N(H_j) p_j(\theta, N)$  coincide with accuracy up to second derivative inclusive.

The approximation of  $p(\theta, N)$  by means of functions of the type (7.2.36) reduces the computation of  $\mathcal{I}[\Delta\mathcal{E}, p(\theta, N)]$  to the problem considered in the preceding section.

In this manner the mean increment of information for an observation taken at the point  $x$  with the sum of the weights  $w$  is equal to [60]

$$\mathcal{I}[\mathcal{E}(\Delta N), p(\theta, N)]$$

$$\simeq \frac{1}{2} \left\{ \sum_{j=1}^v p_N(H_j) d_j(x, N) + \sum_{j=1}^v p_N(H_j) [\eta(x, \hat{\theta}_j) - \sum_{k=1}^v p_N(H_k) \eta(x, \hat{\theta}_k)]^2 \right\} w, \quad (7.2.38)$$

where  $d_j(x, N) = f_j'(x) D_j(N) f_j(x)$ , and the remaining notation is the same as in the follow up to (7.2.36).

VII. Using the results of Theorem 7.2.4 it is not difficult to construct a sequential procedure for designing the experiment, which consists of the following.

After the observations at the  $N$ th stage, the function  $p(\theta_j, N)$  and the quantity

$$\max_x \mathcal{I}[\mathcal{E}(\Delta N), p(\theta_j, N)]$$

are determined.

At the point  $x_N$ , corresponding to the indicated maximum, an observation is taken (or a group of observations). After taking the observation, the function  $p(\theta_j, N+1)$  is determined and all operations are repeated again. Observation is continued as long as the necessary accuracy is not attained. That is, with sufficiently small probability, one of the hypotheses will be separated out and the quantity  $\max_x d_j(x, N)$  [or  $|D_j(N)|$ ] corresponding to it will not be less than some positive number given beforehand.

The computations for the sequential designs are particularly simple in the cases considered in Parts V and VI. The two indicated cases and their obvious combinations may be used in the majority of real situations (these approximations are especially useful for large  $N$ ; this is usually equivalent to the requirement that each distinct observation be not too expensive). In these situations (7.2.28) is replaced by a formula of the type (7.2.35) and the computation is reduced to finding the estimates  $\hat{\theta}$ , their dispersion matrix  $D(N)$ , and the

maximum in  $x$  of some function of the given quantities [cf (7.2.35) and (7.2.38)]

It is interesting to consider the case when one of the hypotheses holds and equation (1.4.4) has a unique root. Under these assumptions, (7.2.35) or (7.2.38) takes on the form

$$\mathcal{J}[\mathcal{E}(\Delta N), p(\theta, N)] \simeq \frac{1}{2} d(x, N) \lambda(x) \Delta N, \quad (7.2.39)$$

and the sequential procedure of designing the experiment considered above coincides with the sequential procedure considered in Chapter 4.

The given coincidence is not accidental. It is a corollary of the fact that

$$\begin{aligned} \Delta I[\mathcal{E}(\Delta N), p(\theta, N)] &= \int p(\theta, N) \ln p(\theta | N) d\theta - \int p(\theta) \ln p(\theta) d\theta \\ &= \int -\frac{1}{2}(\theta - \bar{\theta}) D^{-1}(N)(\bar{\theta} - \theta)(2\pi)^{-m/2} | D(N)|^{-1/2} \\ &\quad \times \exp[-\frac{1}{2}(\theta - \bar{\theta}) D^{-1}(N)(\bar{\theta} - \theta)] \\ &\quad + \int [\ln | D(N) |^{-1/2}] (2\pi)^{-m/2} | D(N) |^{-1/2} \\ &\quad \times \exp[-\frac{1}{2}(\theta - \bar{\theta}) D^{-1}(N)(\bar{\theta} - \theta)] + C_1 \\ &= \ln | D(N) |^{-1/2} + C_2 \end{aligned} \quad (7.2.40)$$

where  $C_1$  and  $C_2$  are constants not depending on  $N$  and the results of the observations.

From (7.2.40) it follows that seeking the maximum of  $\Delta I$  is equivalent to seeking the minimum of the determinant of the matrix  $D(N)$ . Using (7.2.40) and (4.2.8), we can directly obtain (7.2.39).

If there are more than one hypotheses and these are distinct, then as  $N \rightarrow \infty$  the mean increment of information after  $\Delta N$  observations is equal to

$$\mathcal{J}[\mathcal{E}(\Delta N), p(\theta_j, N)] \simeq \frac{1}{2} d_{j_0}(x, N) \lambda(x) \Delta N$$

where  $j_0$  is the true hypothesis [cf (7.2.35) for  $p_N(H_j) \rightarrow 0$  ( $j \neq j_0$ ) and  $p_N(H_{j_0})$ ]. In this manner the sequential design considered in this section is asymptotically optimal in the sense of minimization of the determinant  $|D_{j_0}(N)|$  for the true model.

**EXAMPLE** We continue the example which was introduced in Section 6.4. We will assume that it is necessary not only to clarify which

of the models is true, but also to determine the unknown parameters entering into it as accurately as possible.

Under the assumptions which were made earlier the modeling experiment was conducted according to the sequential design defined by (7.2.35). The results of this experiment are presented in Table 10 and Fig. 36.

Table 10

Results of Experiment Conducted according to the Sequential Design Defined by (7.2.35)

$N$	$x_1$	$T$	$y$	$p_N(H_1)$	$p_N(H_2)$	$p_N(H_3)$	$p_N(H_4)$
1	25	575	0.396				
2	25	475	0.723				
3	125	475	0.422				
4	125	575	0.1297	0.001	0.390	0.612	0.002
5	4.9	600	0.675	0.23( $10^{-3}$ )	0.492	0.505	0.33( $10^{-2}$ )
6	72.7	450	0.706	0.33( $10^{-3}$ )	0.675	0.323	0.12( $10^{-2}$ )
7	150	548	0.085	0.20( $10^{-3}$ )	0.993	0.63( $10^{-2}$ )	0.7( $10^{-7}$ )
8	150	450	0.525	0.17( $10^{-3}$ )	0.995	0.47( $10^{-2}$ )	0.35( $10^{-7}$ )
9	49	600	0.145	0.8( $10^{-5}$ )	0.9998	0.18( $10^{-3}$ )	$\sim 10^{-10}$
10	13	600	0.401	0.7( $10^{-5}$ )	0.9999	0.6( $10^{-4}$ )	$\sim 10^{-11}$
11	39.4	573	0.245	0.5( $10^{-6}$ )	0.9999	0.5( $10^{-5}$ )	$\sim 10^{-14}$
12	150	450	0.544		$\sim 1$		
13	39.0	569	0.279		$\sim 1$		
14	150	450	0.531		$\sim 1$		

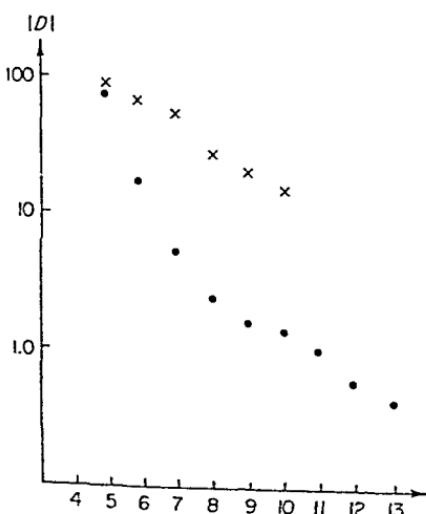


Fig. 36. The value of the determinant of the covariance matrix for the true model obtained for the sequential designs, depending on (6.4.7) (crosses) and (7.2.35) (circles).

Comparison of Tables 9 and 10 shows that for the sequential design carried out based on (7.2.35), the discrimination of the hypotheses at the first few steps is nearly the same. Only after the quantity  $p_N(H_2)$  attains the level 0.90 does the value of this quantity in Table 9 somewhat exceed the corresponding value in Table 10. However, in this case the value of the determinant  $|D_2(N)|$  is essentially smaller than the corresponding values of these quantities, obtained in the experiment, carried out according to the sequential design, aimed only at discriminating hypotheses.

## References

1. V. V. Nalimov and N. A. Chernova. "Statistical Methods of Designing Extremal Experiments." Izd-vo "Nauka," Moscow, 1965.
2. C. R. Hicks. "Fundamental Concepts in the Design of Experiments." Holt, New York, 1964.
3. Yu. P. Adler and Yu. V. Granovskii. A survey of applied papers in the design of experiments. Preprint No. 1, LSM, Izd-vo, Moscow State University, Moscow, 1967.
4. V. V. Nalimov (ed.). "New Ideas in the Design of Experiments: A Collection of Articles." Izd-vo "Nauka," Moscow, 1969.
5. N. P. Klepikov and S. N. Sokolov. "Analysis and Design of Experiments by the Method of Maximum Likelihood." Fizmatgiz, Moscow, 1964.
6. F. R. Gantmacher. "The Theory of Matrices." Chelsea, New York, 1959.
7. A. P. Mishina and I. V. Proskuryakov. "Higher Algebra." S.M.B., Izd-vo "Nauka," Moscow, 1965.
8. E. F. Beckenbach and R. Bellman. "Inequalities." Springer-Verlag, Berlin and New York, 1965.
9. M. G. Kendall and A. Stuart. "The Advanced Theory of Statistics," Vol. 2. Griffin, London, 1961.
10. S. S. Wilks. "Mathematical Statistics." Wiley, New York, 1962.
11. H. Cramér. "Mathematical Methods of Statistics." Princeton Univ. Press, Princeton, New Jersey, 1946.
12. E. S. Venttsel. Theory of Probability. Fizmatgiz, Moscow, 1962.
13. C. R. Rao. "Linear Statistical Methods and Their Applications." Wiley, New York, 1965.
14. S. N. Sokolov and I. I. Silin. "Finding the Minimum of Functionals by the Method of Linearization." Preprint, No. I Ya I, D-810, 1961.

- 15 H O Hartley Modified Gauss method for fitting of nonlinear regression functions *Technometrics* 3, 269 (1961)
- 16 V V Fedorov "Analysis of Experiments in the Presence of Errors in the Determination of the Control Variables" Preprint No 2 LSM, Izd vo, Moscow State University, Moscow, 1968
- 17 I S Berezin and N P Zhukov, 'Computational Methods,' Vols 1, 2 Fizmatgiz, Moscow, 1958
- 18 D Blackwell and M A Girshick 'Theory of Games and Statistical Decisions' Wiley, New York, 1954
- 19 J Kiefer 'Optimum Designs in Regression Problems, II' *Ann Math Statist* 32, 298 (1961)
- 20 J Kiefer and J Wolfowitz The equivalence of two extremum problems. *Canad J Math* 12, 363 (1960)
- 21 S Karlin and W Studden Optimal experimental designs *Ann Math Statist* 37, 783 (1966)
- 22 P Guest The spacing of observations in polynomial regression *Ann Math Statist* 29, 294 (1958)
- 23 V I Smirnov 'A Course of Higher Mathematics' Vol 1 Izd vo Tekhnika-teoret, Moscow, 1957
- 24 G Szego Orthogonal Polynomials" American Mathematical Society Colloquium Publications, Vol 23, American Mathematical Society, Providence, Rhode Island, 1959
- 25 I S Gradshteyn and I M Ryzik Table of Integrals, Series, and Products" Academic Press New York 1965
- 26 I Jahnke, R Linde and R Lyosh Special Functions' Izd vo 'Nauka,' Moscow, 1964
- 27 P Hoel Minimax designs in two-dimensional regression *Ann Math Statist* 36, 1097 (1965)
- 28 K J Arrow, L Hurwicz, and H Uzawa (eds) Studies in Linear and Non-Linear Programming Stanford Univ Press, Stanford California, 1958
- 29 D J Uaill Methods of Seeking Extrema Izd vo 'Nauka,' Moscow, 1967
- 30 V V Fedorov and I S Dubova Methods for Constructing Optimal Designs in Regression Experiments 'Preprint No 4 LSM, Izd vo, Moscow State University, Moscow 1968
- 31 H Chernoff Locally optimal designs for estimating parameters *Ann Math Statist* 24 586 (1953)
- 32 P Hoel Optimum designs for polynomial extrapolation *Ann Math Statist* 36, 1483 (1965)
- 33 J Kiefer and J Wolfowitz On a theorem of Hoel and Levine on extrapolation designs *Ann Math Statist* 36, 1627 (1965)
- 34 S N Sokolov and N B Klepikov Theory of Probability and Its Applications 8, 238 (1963)
- 35 S Karlin and W Studden Tchebycheff systems with Applications in Analysis and Statistics" Wiley (Interscience) New York 1966
- 36 V V Fedorov Properties and Methods for Construction of Point Optimal designs in Regression Experiments Preprint No 5, LSM, Izd vo Moscow State University, Moscow, 1963

37. J. Kiefer. Optimum experimental designs with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp.* I, 381–405 (1965).
38. N. P. Bogachev and I. V. Vzorov. Elastic scatter of protons by means of protons at energies of 660 MeV. *Dokl. Akad. Nauk SSSR* 99, 931 (1954).
39. L. S. Azhgirei and N. P. Klepikov *et al.* Phenomenological Analysis of Interactions at 657 MeV. *Ž. Èksper. Teoret. Phys.* 45, 1174 (1963).
40. F. Legar, V. V. Fedorov, and Z. Yanout, Design of experiments in  $n-p$  scattering. *Ya F* 5, 887 (1967).
41. G. Box and W. Hunter. Sequential designs of experiments for nonlinear models. *Proc. IBM Scientific Computing Symp. Statistics*, p. 113, October 1965.
42. A. Paxman and V. V. Fedorov. Design of estimation and discrimination experiments in  $N-N$  scattering. *Ya F* 6, 853 (1967).
43. F. Lagar and V. V. Fedorov. Design of experiments for the choice between variations of phase shifts. *Ya F* 3, 693 (1966).
44. S. N. Sokolov. Continuous design of regression experiments. *Teor. Verojatnost. i Primenen.* 8, 95, 318 (1963).
45. H. Jeffreys. "Theory of Probability." Oxford Univ. Press, London and New York, 1948.
46. B. Clemmer and R. Krutchkoff. The use of empirical Bayes estimates in a linear regression model. *Biometrika* 55, 525 (1968).
47. V. V. Fedorov. Design of experiments in the case of simultaneous measurements of several response surfaces. Preprint No. 3 LSM, Izd-vo Moscow State University, Moscow, 1968.
48. N. Draper and W. Hunter. Design of experiments for parameter estimation in multiresponse situations. *Biometrika* 53, 525 (1966).
49. E. L. Lehmann. "Testing Statistical Hypothesis." Wiley, New York, 1959.
50. T. W. Anderson. "An Introduction to Multivariate Statistical Analysis." Wiley, New York, 1958.
51. A. Hald. "Statistical Theory with Engineering Applications." Wiley, New York, 1952.
52. V. V. Fedorov. Design of experiments in distinguishing hypotheses of curves by the method of ratio probability. *Zavodskaya Laboratoriya* 34, 314 (1968).
53. J. Kiefer and J. Sacks. Asymptotical optimum sequential inference and design. *Ann. Math. Statist.* 34, 705 (1963).
54. G. S. Fedorov and I. V. Igonina. The information approach to the design of regression experiments. *Zavodskaya Laboratoriya* (in press).
55. W. Hill, W. Hunter, and D. Wichern. A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics* 10, 145 (1968).
56. D. Lindley. On a measure of the information provided by an experiment. *Ann. Math. Statist.* 27, 986 (1956).
57. G. H. Hardy, J. E. Littlewood, and G. Polya. "Inequalities." Cambridge Univ. Press, New York and London, 1959.
58. C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.* 27, 379, 603 (1948).
59. V. V. Fedorov and A. Pazman. Design of physical experiments (statistical methods). *Fortschr. Physik.* 24, 325 (1968).
60. V. V. Fedorov and A. Pazman. Design of Experiments Based on the Measure of Information. Preprint JINR, E5-3249, 1967.

- 61 V. V. Fedorov Sequential Methods for design of experiments in the study of the mechanism of a phenomenon In 'New Ideas in Design of Experiments" (V. V. Nalimov, ed.) Izd vo "Nauka," Moscow, 1969
- 62 G. Eicker Asymptotic normality and consistency of the least squares estimators for families of linear regressions *Ann. Math. Statist.* 34, 447 (1963)

## **Subject Index**

### **A**

- A*-optimal design, 63, 138, 141, 199  
*A*-optimal discrete designs, 169  
A posteriori density, 268  
A posteriori probability  $p(H_i)$ , 258, 267  
A priori probability, 229, 246, 257, 277  
Absolute maxima, 108  
Admissible function, 76, 80  
Allocation, 81, 229  
Asymptotic optimality, 189  
Asymptotically linear-optimal, 184  
Asymptotically normal, 33  
Asymptotically optimal strategies, 179

### **B**

- Bayes approach, 231  
Bayes formula, 237, 258  
Bayes risk, 233, 234, 253  
“Best” estimate, 22  
Best linear estimate, 23ff., 51, 52, 200,  
    210, 233, 237, 250  
Best linear estimator, 27, 50  
    unbiased, 24, 27

- Best quasi-linear estimates, 33, 120, 235  
Binet-Cauchy formula, 15  
Boundary point, 65

### **C**

- Caratheodory’s theorem, 66, 67, 79, 211  
Central limit theorem, 232  
Central moments, 234  
Characteristic numbers, 18  
Chebyshev polynomial, 149  
Chebyshev system, 85ff., 90, 148  
Chebyshev’s inequality, 48  
Closed set, 65  
Comparing designs, 159  
Comparison of experiments, 51ff.  
Complete collection of hypotheses, 228  
Concavity, 213  
Concurring hypotheses, 230  
Conditional density, 237  
Consistent, 31, 45  
Consistent estimator, 33  
Constant efficiency function, 205  
Continuous *D*-optimal designs, 242  
Continuous measure, 62

Continuous normalized designs, 62ff., 156, 204  
 Linear-optimal, 184  
 Contour diagram, 197  
 Control variables, 4, 5, 54 81  
     errors in determination of, 202  
 Convex, 66, 210, 274  
 Convex function, 70, 123  
 Convex hull, 66, 211  
 Convex set, 65, 67, 211  
 Covariance matrix, 176ff

**D**

*D*-criteria, 188  
*D*-optimal design, 63, 68ff., 97, 98, 120, 138, 161, 170, 214, 216  
     continuous, 79  
     discrete, 162, 163  
 Decision rule 228 230, 252 253  
 Design of experiment 58ff  
 Design process 176  
 Determinant, 15 83ff  
     of information matrix, 162  
 Diagonal information matrix, 94  
 Dirac function 62  
 Discrete designs, 155  
 Discrete optimal designs, 160  
     linear 167, 168  
 Discrete spectrum, 81  
 Discriminating experiment, 226ff  
 Discriminating hypotheses, 251, 266, 268  
 Discrimination of hypotheses, 245  
 Dispersion matrix 23, 24 32, 35, 50,  
     *see also* Covariance matrix  
 Distribution of allocation 204, 205  
 Domain of actions, 5

**E**

Efficiency, 254  
     of experiment, 36, 39  
 Efficiency function, 66, 88, 177, 199, 201, 202, 203, 204, 228  
 Ellipsoid of concentration, 54  
 Entropy, 257, 272  
 Entropy measure of information, 268, 270  
 Equivalence theorem, 71, 102

Equiweighted design, 217  
 Errors in control variables, 40, 42  
 Estimate, 21  
     consistent, 22  
     efficient, 22, 33  
     sufficient, 22  
     unbiased, 22  
 Estimation of dispersion, 36  
 Estimator, 35  
 Expectation operator, 21, 153, 233  
 Expected increment of information, 270, 271, 276  
 Expected information, 274  
 Experiment, 52ff., 252  
 Experimental analysis, 21  
 Extrapolation at point, 145  
 Extremal points, 280  
 Extremal problems 60, 79, 214

**F**

Factor space, 5 230, 237 241, 242, 245  
 Fisher information matrix 27, 120  
 Frobenius formula, 16 129  
 Functional, 143  
 Functional  $\mathcal{L}(\delta)$  57

**G**

Generalized dispersion, 27  
 Generalized likelihood function, 250  
 Generalized likelihood function ratio, 249  
 Generalized loss, 264

**H**

Hadamard inequality, 19  
 Hermite polynomial, 89  
 Homogeneity of experimental conditions, 39  
 Hypercube, 76  
 Hypersurface, 75

**I**

Incomplete collection of hypotheses, 230  
 Incorrect hypothesis, averaged possible losses for accepting, 241

Increment of information, 259  
 Indifference zone, 229  
 Information matrix, 59, 65ff., 74, 77, 79,  
     98, 102, 155, 156, 161, 205, 210  
 Initial approximation, 35  
 Interior point, 65  
 Invariance property, 80  
 Invariant, 81  
 Inverse of matrix, 14  
 Inversion  
     of information matrix, 104, 107  
     of matrix, 174, 176  
 Iterative procedures, 99, 102, 112, 135,  
     136, 169, 220, 223, 248  
 Iterative process, 35, 45, 104, 216  
     A-optimal designs, 140  
     Q-optimal designs, 144

**J**

Jacobi polynomial, 89, 91

**L**

Lagrange interpolation polynomial, 147  
 Laguerre polynomial, 89  
 Large-dimensional experimental data, 30  
 Legendre polynomial, 89  
 Limiting design  $\tilde{\epsilon}$ , 137  
 Linear optimal, 136  
 Linear-optimal design, 122, 125, 126,  
     170, 221  
     spectrum, 128  
 Linear unbiased estimators, 25  
 Locally  $D$ -optimal design, 197, 245  
 Locally optimal statistical designs, 186  
 Loss function, 56, 230, 231

**M**

Matrix  
     algebra, 12  
     characteristic polynomial, 18  
     differentiation, 20  
     integration, 20  
     minors, 16  
     partitioned, 16  
     positive definite, 17ff.

positive semidefinite, 18, 19  
 rank, 15, 16, 20  
 singular, 15  
 symmetric, 17  
 Maximum likelihood, 33  
 Minimum risk, 229  
 Mean increment of information, 272, 282  
 Mechanism of phenomena, 39  
 Metric, 279  
 Minimax criterion, 138  
 Minimax design, 63, 68  
     in space of parameters, 63  
 Minimum of quadratic form, 44  
 Minor matrices, 84

**N**

Neutron-proton scatter, 253  
 Nondegenerate design, 223, 252  
 Nonlinear parametrization, 33, 35, 189,  
     249, 263, 267  
     of response surface, 186, 188  
 Nonnormalized designs, 184  
     discrete, 161  
 Nonsequential experimental design, 247  
 Normal distribution function, 48  
 Normalized continuous designs, 210  
 Normalized design, 59, 67, 184  
     discrete, 156  
 Normalized  $D$ -optimal design, 178  
 Numerical construction  
     of  $D$ -optimal designs, 99, 114  
     of optimal linear designs, 169  
     of truncated  $D$ -optimal designs, 119,  
         133, 184  
 Numerical method, 223

**O**

One point designs, 275  
 Open set, 65  
 Operator, 158  
 Operator  $L$ , 167, 168  
 Optimal decision rule  $\delta$ , 231, 253  
 Optimal design, 7, 74, 79, 146  
 Optimal rule, 229  
 Optimum decision, 231  
 Orthogonal polynomials, 91, 92

**P**

- Polya's theorem 87  
 Polynomial 86 129  
 Polynomial regression 85  
 Positive semidefinite matrix 27 138, 222  
 Posterior distribution 237  
 Power functions 85  
 Preliminary experiment 8 179 187 188  
 Priming experiment 252  
 Principal minors 18  
 Prior information 191 192 231  
 Product of matrix 13  
 Properties of estimator statistical 44

**Q**

- Q*-optimal designs 142 143  
 Quadratic form 32 45 66 75 111  
 Quadratic loss 153  
 Quasi linear 200

**R**

- Rank 67  
 Region of possible measurements 5  
 Regression analysis 21  
 Regression problem 82 218 219  
 Regression surfaces 166  
 Response surface 59 69 76 175  
 Risk function 196  
 Robust 23  
 Rounded-off design 110

**S**

- Search for optimal designs 75  
 Sequential construction 245  
 Sequential design 7 172ff 224 263 283  
 Sequential estimation 262  
 Sequential linear optimal designs 183  
 Sequential procedure 281 282  
 Several random quantities 209

- Several response surfaces 221  
 Simultaneous measurements 209 224  
 Simultaneous observations 49  
 Singular information matrices 118  
 Singular matrix 125  
 Spectrum 61 67 72 76 111 162 163  
 164 210  
 of design 59  
 Sphere 65  
 Statistical design 7  
 Strategy 179  
 Strong law of large numbers 46  
 Strongly biased 40  
 Submatrix of matrix 13  
 Superposition of normal distributions 250

**T**

- Taylor series 41 260  
 Transpose matrix 13  
 Trigonometric regression 95 96 97  
 Truncated *D*-criteria 188 221  
 Truncated *D*-optimal designs 114 115  
 118 119 178

**U**

- Unbiased 32  
 Unbiased estimate 37 39  
 Uniform design 92 218 219  
 Uniform spectrum 92 96  
 Uniqueness of information matrix 128

**W**

- Weight function 246 266  
 Weight matrix 209  
 Weight multiplier 265 278  
 Weights 109 235  
 Weighted sum 69  
 of squares 30