Computational Aspects of Discrimination between Models of Dynamic Systems

Faculty of Electrical Engineering, Computer Science and Telecommunications
University of Zielona Góra

Lecture Notes in Control and Computer Science
Volume 10

Bartosz Kuczewski

Computational Aspects of Discrimination between Models of Dynamic Systems

University of Zielona Góra Press, Poland
2006

Bartosz Kuczewski
Institute of Control and Computation Engineering
University of Zielona Góra
ul. Podgórna 50
65-246 Zielona Góra, Poland
e-mail: B.Kuczewski@issi.uz.zgora.pl

Referees:

- Andrzej KRÓLIKOWSKI, Poznań University of Technology
- Krzysztof GAŁKOWSKI, University of Zielona Góra

The text of this book was prepared based on the author's Ph.D. dissertation

The work was supported by the State Committee for Scientific Research (KBN)
in Poland, grant No. 3 T11C 035 27

ISBN 83-7481-030-0

Camera-ready copy prepared from the author's LaTeX2$_\varepsilon$ files.
Printed and bound by University of Zielona Góra Press, Poland.

Contents

Acknowledgments

At this point, I would like to express my gratitude to the supervisor of this work, Professor Dariusz Uciński. Without his continuous support, fresh ideas and concepts and, last but not least, patience, this dissertation would have never been completed. His contribution to the final form of the dissertation cannot be overestimated.

I also thank my friends Maciej Patan, Bartłomiej Sulikowski and Wojciech Paszke for many enlightening discussions, as well as for support and help regarding editorial technicalities. Przemysław Baranowski is gratefully acknowledged for his help in the field of parallel computing.

Finally, no words exist to express my gratitude to my parents and siblings.

Notation

Symbols

$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
$\mathrm{Sym}(m)$	set of all symmetric $m \times m$ matrices
$\mathrm{NND}(m)$	set of all symmetric, nonnegative definite $m \times m$ matrices
$\mathrm{PD}(m)$	set of all symmetric, positive definite $m \times m$ matrices
t	time
x	spatial coordinate
$\vartheta, \widehat{\vartheta}$	unknown parameter vector and its estimate, respectively
m	dimension of parameter space
d	number of model outputs
$\xi_N(\,\cdot\,)$	exact design of an experiment
$\xi(\,\cdot\,)$	continuous design of an experiment
I_d	$d \times d$ identity matrix
0_d	$d \times d$ zero matrix
$\Delta_1(\,\cdot\,)$	T-optimality criterion
N	total number of measurements
n	number of support points
p	vector of design weights
$\mathcal{M}$	analytical model
ε	measurement error
ϵ	numerical tolerance

Operators and functions

η	model response
E	expectation
$\mathrm{col}[\,\cdot\,, \ldots, \cdot\,]$	column vector
$\mathrm{Diag}[\,\cdot\,, \ldots, \cdot\,]$	diagonal matrix with given diagonal entries
Cov	covariance
Var	variance
$\mathrm{trace}(A)$	trace of matrix A
$\det(A)$	determinant of matrix A
$\|\,\cdot\,\|$	Euclidean norm
$\mathrm{supp}\,\xi$	support of a measure ξ
$\phi(\,\cdot\,)$	variance prediction function
$\psi(\,\cdot\,)$	sensitivity function

Given a function g defined on a set Y, we write

$$\operatorname{Arg}\min_{y\in Y} g(y) = \left\{ y \in Y : g(y) = \min_{y\in Y} g(y) \right\},$$

i.e., $\operatorname{Arg}\min_{y\in Y} g(y)$ consists of all the y's that attain the minimum of g over Y. We will use the symbol

$$\arg\min_{y\in Y} g(y)$$

to denote any element of $\operatorname{Arg}\min_{y\in Y} g(y)$. The same convention will apply to maxima.

Abbreviations

ARS	Adaptive Random Search
DPS	Distributed Parameter System
FIM	Fisher Information Matrix
MCMC	Monte Carlo Markov Chain
MPI	Message Passing Interface
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
SIP	Semi-Infinite Programming
SMP	Symmetric Multi-Processing
SQP	Sequential Quadratic Programming
W-F	Wynn-Fedorov

Chapter 1

INTRODUCTION

There are many situations where accurate statistical models of real systems are desirable. For example, engineers often require models for prediction or control purposes (Goodwin and Payne, 1977; Żak, 2003; Bryson, 1999; Bryson, 2002; Burl, 1999; Bertsekas, 2000; Banks and Kunisch, 1989; Banks *et al.*, 1996; Królikowski, 2004). Other areas of applications include chemistry, pharmaceutics, ecology, economics, biology and sociology, cf. e.g., the monographs by (Schittkowski, 2002; Burnham and Anderson, 2002; Pilling and Seakins, 1995).

Mechanistic models can often be obtained from physical reasoning. But if a process is not fully known, there may be several plausible models. Then a properly designed experiment can help in discriminating between the competing models. However, the ability to obtain an accurate model is limited by the presence of random fluctuations such as unmeasurable disturbances and measurement errors. Hence, careful consideration of the error variability while designing the experiment is crucial for model discrimination. Experimental designs for measurements leading to efficient and precise estimation of the parameters and also for discrimination between models have been developed for simple models. An introduction is in Chapters 18–20 of (Atkinson and Donev, 1992), with a fuller survey by (Atkinson, 1992).

For dynamic systems, the design of the experimental conditions so that the experiment is maximally informative includes a choice of input and measurement ports, test signals, sampling instants, presampling filters, and some parameters such as temperature, pressure, etc., cf. (Goodwin and Payne, 1977; Królikowski and Eykhoff, 1985; van de Wal and de Jager, 2001; Uciński, 2005; Patan, 2004; Spall, 2003). Although a substantial literature has accumulated on the theory of optimum experimental design and numerous results have been applied with great success in practice over the last three decades (Atkinson and Donev, 1992; Fedorov, 1972; Fedorov and Hackl, 1997; Kiefer and Wolfowitz, 1959; Pázman, 1986; Pukelsheim, 1993; Rafajłowicz, 1996; Walter and Pronzato, 1997; Müller, 1998; Schwabe, 1996; Silvey, 1980; Cox and Reid, 2000; Zarrop, 1979; Antony, 2003), much still has to be done, especially in the context of discriminating among different candidate models of dynamic systems (especially systems described by ordinary or partial differential equations) for multivariate data, where there are few useful results. The problem arises when there are two or more rival models and the purpose of the experiment may be to determine which, if any, of the models are adequate. The practical potential of the corresponding results could hardly be overestimated.

Various design criteria for discrimination between models were considered by (Box and Hill, 1967; Atkinson and Fedorov, 1975a; Atkinson and Fedorov, 1975b; Fedorov and Khabarov, 1986; Ponce de Leon and Atkinson, 1991; Felsenstein, 1992; Burke *et al.*, 1994; Müller and Ponce de Leon, 1996; Stewart *et al.*, 1998). The criterion, called T-optimality, introduced by (Atkinson and Fedorov, 1975a) and then extended in (Atkinson and Fedorov, 1975b) in a single response case, has attracted our attention as it has an interesting statistical interpretation as the power of a test for the fit of a second model when the first one is true.

In this dissertation we consider T-optimum designs for discrimination between two and more rival multi-response models with observations corrupted by normally distributed noise with zero mean and a known covariance matrix which may depend on time and/or unknown parameters. We assume that the observations are not correlated in time, but we admit of correlations among different responses. The T-optimality criterion was generalized for this case in (Uciński and Bogacka, 2005; Uciński and Bogacka, 2004) and its use leads to solving a maximin problem. In the context of chemical kinetics considered therein, the design factors are the sampling strategy and experimental conditions such as temperature or the initial concentrations of reactants. Both affect the accuracy of model discrimination. Clearly, the inclusion of a variety of experimental conditions complicates both the mathematical and computational problem of finding an optimum design. However, their inclusion is very important in applications.

A further difficulty is that the model functions may be given only implicitly, as solutions, usually numerical, to a system of ordinary, differential-algebraic or partial differential equations. To overcome all these difficulties, necessary and sufficient conditions for optimality are usually formulated. These constitute a generalization of the well-known Equivalence Theorem (Atkinson and Fedorov, 1975a; Atkinson and Fedorov, 1975b; Atkinson and Donev, 1992; Fedorov and Hackl, 1997). They can be applied when checking potentially optimal designs.

Identification of the correct dynamic model is a process consisting of several stages, which include data collection, fitting and diagnostic checking. The fitting and checking may not lead to a unique conclusion if the observations are taken in parts of the regression region and under the experimental conditions which suit more than one model. (Box and Hill, 1967) illustrate this problem for a catalytic reaction which may be consistent with four different mechanisms. They say "Unfortunately, it is easy to collect data that are well fitted by a large number of different models. Different research groups commonly claim widely varying mechanisms for the same mechanical system". Hence, a planning stage should precede the data collection. For each specific dynamic system the data collection needs to be especially designed (tailor-made). It is not really possible to find a common design which would suit a wide range of dynamic systems, though, procedures based on general theoretical results can be proposed for calculating the designs for model discrimination for some classes of models.

The results presented in this dissertation are in general form and they can be directly applied to various deterministic dynamic systems which constitute our primary motivation to study T-optimal designs. These systems are continuous-

time models and we are specifically interested in their largest class given in terms of ordinary differential equations. To be specific, we consider two competing models described by the following (possibly nonlinear) equations:

$$\mathcal{M}_\ell : \quad \frac{\mathrm{d}v_\ell(t)}{\mathrm{d}t} = f_\ell(t, v_\ell(t), \vartheta_\ell), \quad v_\ell(0) = v_0, \quad \ell = 1, 2, \tag{1.1}$$

where t is time, v_ℓ stands for a vector-valued function $v_\ell : T \to \mathbb{R}^d$ (also called the state), $T = [0, t_f]$ for a given t_f, $\vartheta_\ell \in \Theta_\ell$ denotes a vector of constant unknown parameters, v_0 signifies the vector of initial state values, and known functions f_ℓ are required to be continuous, Θ_ℓ being a given set. Note that (1.1) defines implicitly mappings $\eta_\ell : T \times \Theta_\ell \to \mathbb{R}^d$ such that $\eta_\ell(\,\cdot\,, \vartheta_\ell)$ coincides with the solution $v_\ell(\,\cdot\,)$ for any fixed ϑ_ℓ, $\ell = 1, 2$.

The solutions of mechanistic models (1.1) give the expectations of the two competing statistical models, where random errors of observations are included. Discrimination between models $\mathcal{M}_1$ and $\mathcal{M}_2$ based on an experimental design is equivalent to discrimination between two statistical models (Uciński and Bogacka, 2005):

$$\mathcal{S}_\ell : \quad y_{ij} = \eta_\ell(t_i, \vartheta_\ell^{(0)}) + \varepsilon_{ij}, \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, n, \quad \ell = 1, 2, \tag{1.2}$$

where the observations $y_{ij} \in \mathbb{R}^d$ are taken at discrete times $t_1, \ldots, t_n$, η_ℓ being a solution of the differential equations for some prior values $\vartheta_\ell^{(0)}$ of the parameters ϑ_ℓ, r_i denoting replications of the observations. This leads to the question of how to select an appropriate sampling strategy. A serious complication here is the presence of unknown parameter vectors ϑ_ℓ. They have to be estimated (if not given a-priori) and one of several possible approaches is to design an experiment for both estimation of ϑ_ℓ and discrimination between the models (Atkinson, 1992; O'Brien and Rawlings, 1996). In this dissertation we are interested in model discrimination where one of the models is assumed to be known completely and the parameters of the other model are to be estimated. Another complication is that the relationship η_ℓ is defined implicitly through the solution of the ordinary differential equation, but this obstacle is only of minor importance due to the availability of extremely efficient solvers performing numerical integration of vector differential equations.

For example, chemists often deal with multi-response data in experimental studies of chemical reactions. An important special class of the corresponding nonlinear models is that in which the responses are described by a system of ordinary differential equations. These models are used very frequently in chemical kinetics (Bates and Watts, 1988) and constitute a wide class of dynamic systems. Parameter estimation methods for such chemical kinetic models are relatively well-developed (Bard, 1974; Seber and Wild, 1989; Schittkowski, 2002). But, as in other dynamical systems, here too, several alternative models are often proposed for the same physical phenomenon. We then wish to conduct experiments that would enable us to select the 'best' model, i.e., the one that best fits the data. Each one of these models implicitly attempts to predict the responses as functions of time and parameters.

As an example of this type, consider two chemical reactions: a reaction where substance A changes into substance B, which in turn changes into substance C, but the first part of the reaction may be reversible, that is

$$A \underset{k_3^{(1)}}{\overset{k_1^{(1)}}{\rightleftharpoons}} B \overset{k_2^{(1)}}{\rightarrow} C, \tag{1.3}$$

with rate constants $k_1^{(1)}$, $k_2^{(1)}$ and $k_3^{(1)}$ (for the rate of the reverse reaction) and a very similar, but irreversible process

$$A \overset{k_1^{(2)}}{\rightarrow} B \overset{k_2^{(2)}}{\rightarrow} C, \tag{1.4}$$

with rates $k_1^{(2)}$ and $k_2^{(2)}$.

The first part of (1.3) is an opposing reaction where the reverse reaction may often be questionable. Enzyme catalysis provides examples of such complex reactions where an enzyme (E) forms a complex (ES) with a substrate (S) which then reacts to form product (P) and also regenerates the enzyme:

$$E + S \rightleftharpoons ES \rightarrow P + E, \tag{1.5}$$

see Chapter 8 of (Pilling and Seakins, 1995). In many chemical systems the reverse reaction is often neglected, when it is a very slow one, resulting in a simpler consecutive model (1.4). Whether this a right thing to do may be examined using a model discrimination technique. In this dissertation we provide such a technique for this kind of models. However, the presented results are general and can be applied to other dynamic systems as well.

The concentrations of any, or all, of the reactants can be measured. The changes in concentrations are governed by ordinary differential equations. The first competing model is given implicitly, as a solution of the system

$$\mathcal{M}_1 : \begin{cases} \dfrac{\mathrm{d}[A]}{\mathrm{d}t} = -k_1^{(1)}[A]^{\lambda_1^{(1)}} + k_3^{(1)}[B]^{\lambda_3^{(1)}}, & [A]_{t=0} = a_0, \\ \dfrac{\mathrm{d}[B]}{\mathrm{d}t} = k_1^{(1)}[A]^{\lambda_1^{(1)}} - k_2^{(1)}[B]^{\lambda_2^{(1)}} - k_3^{(1)}[B]^{\lambda_3^{(1)}}, & [B]_{t=0} = b_0, \\ \dfrac{\mathrm{d}[C]}{\mathrm{d}t} = k_2^{(1)}[B]^{\lambda_2^{(1)}}, & [C]_{t=0} = c_0, \end{cases} \tag{1.6}$$

where $[A]$, $[B]$ and $[C]$ are concentrations of chemical compounds A, B and C as functions of time t, and a_0, b_0 and c_0 stand for initial reactant concentrations. The $\lambda_i^{(1)}$'s denote the reaction orders. Similarly, the second model is given implicitly by

$$\mathcal{M}_2 : \begin{cases} \dfrac{\mathrm{d}[A]}{\mathrm{d}t} = -k_1^{(2)}[A]^{\lambda_1^{(2)}}, & [A]_{t=0} = a_0, \\ \dfrac{\mathrm{d}[B]}{\mathrm{d}t} = k_1^{(2)}[A]^{\lambda_1^{(2)}} - k_2^{(2)}[B]^{\lambda_2^{(2)}}, & [B]_{t=0} = b_0, \\ \dfrac{\mathrm{d}[C]}{\mathrm{d}t} = k_2^{(2)}[B]^{\lambda_2^{(2)}}, & [C]_{t=0} = c_0, \end{cases} \tag{1.7}$$

The parameters $k_i^{(\ell)}$ and $\lambda_i^{(\ell)}$ are unknown and they are estimated from the experimental data. Each of the models (1.6) and (1.7) covers a wide class of chemical reactions. This is due to allowing for any orders $\lambda_i^{(\ell)}$. The systems of differential equations in such a case do not have analytical solutions and, to our knowledge, they have received little attention in the statistical literature. Designs for estimation of the model parameters were calculated for the second model by (Atkinson and Bogacka, 2002) and some approach to discriminate between $\mathcal{M}_1$ and $\mathcal{M}_2$ was outlined in (Uciński and Bogacka, 2005).

Suppose that we already have estimates $\widehat{\vartheta}_1 = (\widehat{k}_1^{(1)}, \widehat{k}_2^{(1)}, \widehat{k}_3^{(1)}, \widehat{\lambda}_1^{(1)}, \widehat{\lambda}_2^{(1)}, \widehat{\lambda}_3^{(1)})^{\mathrm{T}}$ (obtained e.g. from a preliminary experiment) for the respective parameters $k_i^{(1)}$ and $\lambda_i^{(1)}$, $i = 1, 2, 3$, in model (1.6). We argue that model $\mathcal{M}_1$ in (1.6) with the estimated parameters is correct, but at the same time we wish to test this assertion against the one saying that the alternative model $\mathcal{M}_2$ in (1.7) is true, where the vector of parameters $\vartheta_2 = (k_1^{(2)}, k_2^{(2)}, \lambda_1^{(2)}\lambda_2^{(2)})^{\mathrm{T}}$ belongs to a known set Θ_2.

Clearly, such model discrimination should be based on additional experimental observations. The experiment consists therefore in measuring the concentrations of A, B and C after the reaction had been running for a time t. One experimental run yields one three-element observation vector y_i (measurements of concentrations $[A]$, $[B]$ and $[C]$) and the experimental design is a list of N times t_i, $i = 1, \ldots, N$, $N = \sum_{i=1}^{n} r_i$, not necessarily distinct, at which measurements are to be made. The aim at this point is to discriminate between the competing models. The simple model is often preferred even though a more complex model is bound to give a better fit to the data. This preference is tied up with some vague notion of model cost effectiveness, but this may be difficult to quantify in practical situations. The problem is compounded by the fact that it will be seldom, if ever, the case that the true mechanism corresponds to any of the models under study (Goodwin and Payne, 1977). Here we adopt an approach based on the T-optimality criterion which has been used since the mid-seventies for single-response models (Atkinson and Fedorov, 1975a; Atkinson and Fedorov, 1975b). A working assumption in this method is that one of the competing models, $\mathcal{M}_1$ or $\mathcal{M}_2$, is the true model, its number and the corresponding parameters being all known. Although this may seem to be a somewhat abstract problem, the properties of the designs thus obtained are useful in real situations, as indicated by (Atkinson and Fedorov, 1975a), when neither the true model nor its parameter values are known.

Consequently, with no loss of generality, we can assume that $\mathcal{M}_1$ is true, i.e., that the true response $\eta(\,\cdot\,)$ coincides with $\eta_1(\,\cdot\,, \widetilde{\vartheta}_1)$, where $\widetilde{\vartheta}_1$ is regarded as known before the experiment (this value could be obtained on the basis of some preliminary experiment). Then, to design an optimal experiment for discrimination between $\mathcal{M}_1$ and $\mathcal{M}_2$ means to select the measurement points $t_1, \ldots, t_n$ so as to maximize the minimum of the sum of squares for the lack of fit of the second model:

$$\Delta_1 = \min_{\vartheta_2 \in \Theta_2} \sum_{i=1}^{n} p_i \|\eta(t_i) - \eta_2(t_i, \vartheta_2)\|^2, \tag{1.8}$$

where $p_i = r_i/N$, $i = 1, \ldots, n$. In a linear model, the power of the F-test for

departures from the true model is an increasing function of Δ_1 (Atkinson and Fedorov, 1975a). Furthermore for normally distributed errors and large N, Δ_1 is proportional (in mean) to the ratio of the likelihoods associated with both the models.

In spite of the seemingly simple and intuitively clear form of the criterion (1.8), its efficient numerical maximization has still been remaining an insoluble problem, which discourages researchers from attempts to use T-optimum experimental designs in practice. This is surprising and an even more confounded nuisance today, when new software and hardware tools provide capabilities for intricate analysis of many difficult performance aspects of systems. These analysis models can be coupled with numerical optimization software to generate better designs iteratively. Such tools have dramatically increased in sophistication, and engineers are called to cope with highly complex problems. This was a main stimulus to prepare this dissertation.

Our main goal was to develop the background needed to solve computational problems and to provide efficient numerical methods of constructing T-optimum designs for dynamic processes described by ordinary and partial differential equations.

The success of such an attempt depends strongly on how well the design problem has been formulated for optimization study, and on how familiar the designer is with the workings and pitfalls of iterative optimization techniques. Raw computing power is unlikely to ease this burden of knowledge.

A starting point for this project was the work by (Atkinson and Fedorov, 1975a) who proposed an algorithm for generating approximations to T-optimum designs, which has remained since then the only known computational tool in this context. But the major drawback of the method was the lack of its convergence analysis. In fact, the method, as it was formulated, was not globally convergent, cf. (Fedorov and Hackl, 1997). The obvious task of the present research was thus to look closely at Fedorov's algorithm. This resulted in the formulation of a family of methods which combine some features of the original Fedorov method and, at the same time, possess global convergence properties.

The work outlined in this dissertation is intended to develop the underlying theory and to construct efficient numerical procedures for determining optimal experimental designs for discriminating between several rival multiple response nonlinear models. The proposed solutions have been tested on practical process engineering examples, thereby indicating their potential applications in numerous disciplines. The following is a concise list of detailed objectives:

1. The most difficult aspect of the structure discrimination problem is to obtain a meaningful criterion. Thus the very first objective is to introduce appropriate optimality criteria and characterize the corresponding optimal solutions. Atkinson and Fedorov (1975) introduced designs for discriminating between single response models and (Uciński and Bogacka, 2005) generalized those results to the multiresponse framework. In our programme we will further extend this work to various application-driven situations, e.g., design for distributed parameter systems or for fault detection in dynamic processes.

2. We are also interested in providing sequential design algorithms. Initially, some known algorithms, which are efficient in the single response case, will be generalized to the multiple response case and to dynamic systems. In addition, recent advances in semi-infinite optimization will be exploited in order to make the computation process much more efficient. What is more, a flexible relaxation algorithm RATO will be proposed and its global convergence in a finite number of steps will be proven. Such an approach constitutes a major novelty which has not been considered yet in optimum experimental design.

3. In order to test the proposed methods, various frameworks and applications to process engineering examples will be considered, thus validating the practical usefulness of the proposed approach. The problems to be considered include optimization of observation policies regarding chemical kinetics of reactions, optimization of air quality monitoring networks, and fault detection and isolation for distributed parameter systems.

4. Based on methods elaborated in the dissertation, several portable computer programs and software packages have been developed in Fortran 95 using the IMSL numerical library and MATLAB for calculation of optimum designs for selected classes of dynamic systems. These serve not only to demonstrate the algorithms and the underlying principles but also they might be used as a basis for implementation of expert systems supporting experimenters dealing with complex environmental and industrial processes.

The outline of the dissertation structure is as follows:

In Chapter 2 we give a brief introduction to modern optimum experimental design. The presented material is oriented so as to make the reader familiar with basic notions and technicalities developed within the framework of optimum design for parameter estimation.

In Chapter 3 we present a general set-up of the multi-response model discrimination problem and the necessary and sufficient conditions for the optimality of a design, i.e., the so-called equivalence theorem. The T-optimality criterion is defined and its properties are exhaustively studied. The usefulness of the incorporated equivalence theorem is presented on a few examples when analytical solutions can be found.

In Chapter 4 we present our main result, namely the numerical algorithms developed to solve the T-optimum design problem. Particularly, a novel relaxation algorithm RATO is presented and its convergence in a finite number of steps is proved. Moreover, a thorough analysis of additional numerical problems associated with the RATO scheme is presented, including a proposition of appropriate regularization for non-smooth functions being optimized. Also, the possibility of using Semi-Infinite Programming to obtain T-optimum designs is delineated. Several numerical examples are presented to illustrate the effectiveness of the outlined approaches.

Chapter 5 deals with some additional extensions to the basic T-optimum design problem. In particular, the convergence of the Wynn-Fedorov type algorithm

is studied for the newly proposed DT-optimum criterion which provides a reasonable balance between model discrimination and parameter estimation. What is more, the problem of T-optimum designs on finite design spaces is described including numerical algorithms to solve it. Also the adaptation of Rafajłowicz's selective random search technique for T-optimum designs is presented. Finally, a way to parallelize the Wynn-Fedorov algorithm is outlined which constitutes a very promising result in the context of applications.

Chapter 6 presents some systematic approaches based on T-optimum experimental design and oriented towards applications. First, the problem of discrimination between several competing models is investigated. Then, a very useful concept of the replication-free designs is considered in the context of discrimination between models. Finally, the problem of optimum sensor location for discrimination between models of distributed parameter systems is considered, also in the context of fault detection in industrial processes and correlated observation errors.

Conclusions are discussed in Chapter 7.

Chapter 2

AN OVERVIEW OF OPTIMUM EXPERIMENTAL DESIGN THEORY

First problems of the broadly defined optimization of experimental efforts were formulated almost a hundred years ago (Atkinson and Bailey, 2001). Generally, optimum experimental design consists in determination of a "best" schedule of taking observations of the investigated process or phenomenon in the sense of minimizing some performance criterion or maximizing the confidence of detection or discrimination.

As regards historical origins, the developed methods were mainly inspired by real-world optimization problems which can be divided into two groups:

- agricultural or medical experiments,
- industrial experiments.

A typical example of a problem belonging to the former group is a clinical trial, e.g., the problem consisting in comparing the effectiveness of a new drug with placebo. Here special techniques of data processing such as blocking are usually applied (this induces the notion of the so-called block designs) and a "design" is also understood as a manner of handling available data (Bogacka, 1995; Gilmour and Trinca, 2000) (in such a problem the treatment factors usually have both qualitative and quantitative parts, and the resulting optimization problems are of combinatorial nature). That kind of designs is, however, beyond the scope of this dissertation. The reader interested in this area of research can be referred, e.g., to the comprehensive monograph (Bailey, 2004) and the references given therein.

In most cases the methods of the second group are equivalent to optimum design for regression performed for the parameter estimation of an assumed model and originate from the ideas set forth by (Kiefer and Wolfowitz, 1959). Here the design is understood as a collection of variables which completely determine the strategy of taking measurements or observations of the investigated process responses or states.

The main effort of this dissertation is concentrated on application of optimum experimental design methods in structural identification which is understood as discrimination between several competing models. However, to provide the reader with basic ideas and concepts of designing experiments, first we present a brief overview of the state of the art in that field for parameter estimation.

It is justified by the fact that a majority of key ideas are common for both the problems. Since the relevant theory is rather well known and the appropriate literature is abundant (Fedorov, 1972; Silvey, 1980; Pázman, 1986; Atkinson and Donev, 1992; Pukelsheim, 1993; Rafajłowicz, 1996; Fedorov and Hackl, 1997; Müller, 1998; Cox and Reid, 2000; Walter and Pronzato, 1987; Walter and Pronzato, 1997; Rafajłowicz, 2006; Atkinson *et al.*, 2001), including modern applications in the field of parameter estimation of distributed parameter systems (Uciński, 1999; Patan, 2004; Uciński, 2005), we limit ourselves only to basic concepts which will serve as an introduction to further considerations.

2.1. Optimum design for regression

2.1.1. Linear models

For the sake of simplicity, we introduce concepts of convex design theory based on univariate models which are linear with respect to the unknown parameters. Such models can be described by the following equation:

$$\eta(x,\vartheta^\star) = (\vartheta^\star)^{\mathrm{T}} f(x) = \sum_{i=1}^{m} \vartheta_i^\star f_i(x), \tag{2.1}$$

where $\vartheta^\star \in \mathbb{R}^m$ stands for the true value of the vector of unknown but constant parameters and $f(x) = [f_1(x), f_2(x), \ldots, f_m(x)]^{\mathrm{T}}$ denotes the vector of known real-valued basis functions $f_i(x)$, $i = 1, \ldots, m$ which are continuous on a given compact set $X \in \mathbb{R}^s$. Additionally, we impose the restriction that $f_1(x), f_2(x), \ldots, f_m(x)$ must be linearly independent.

The output equation describing the measurements has the form

$$z_i = \eta(x_i,\vartheta^\star) + \varepsilon_i = (\vartheta^\star)^{\mathrm{T}} f(x_i) + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.2}$$

where $x_i \in X$ are the points at which the process response y is observed. The ε_i's constitute additive purely random Gaussian errors, i.e., they are zero mean and independent of each other. The most common assumption about the variances of disturbances ε_i is about their equality, that is to say,

$$\mathrm{Var}(\varepsilon_i) = \sigma^2 \tag{2.3}$$

with a constant, but possibly unknown value of $\sigma > 0$. However, the variances might as well be distinct:

$$\mathrm{Var}(\varepsilon_i) = \sigma^2(x_i) = \sigma^2 \varrho^{-1}(x_i), \tag{2.4}$$

where $\sigma > 0$ is unknown and the sequence of positive values $\varrho(x_i)$ is known and can be interpreted as a relative accuracy of observations at particular points x_i. The x_i's are selected so as to satisfy

$$\mathrm{rank}[f(x_1), f(x_2), \ldots, f(x_N)] = m, \tag{2.5}$$

which actually means that $N \geq m$, i.e., the number of measurements must be at least equal to the number of the estimated parameters.

Equation (2.2) can be alternatively rewritten in matrix form

$$y = F_N \vartheta^\star + \varepsilon, \tag{2.6}$$

where

$$\begin{aligned} y &= [y_1, y_2, \ldots, y_N]^{\mathrm{T}}, \\ \varepsilon &= [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N]^{\mathrm{T}}, \\ F_N &= [f(x_1), f(x_2), \ldots, f(x_N)]^{\mathrm{T}}. \end{aligned}$$

The objective of parameter estimation is to find an estimate $\widehat{\vartheta} \in \Theta_{ad}$, Θ_{ad} being a set of admissible parameter values for the unknown true parameter vector $\vartheta^\star$, based on the set of process observations (2.2) such that the predicted response of the model η is close enough to these process observations (in the sense of a given quality measure). Here it is assumed that parameter estimates $\widehat{\vartheta}$ are obtained by applying the weighted least-squares criterion in the form (Seber and Lee, 2003; Montgomery *et al.*, 2001)

$$\mathcal{J}(\vartheta) = \frac{1}{2} \sum_{i=1}^{N} w_i \left(y_i - \vartheta^{\mathrm{T}} f(x_i) \right)^2, \tag{2.7}$$

where $w_i \geq 0$ constitute weighting coefficients, which yields

$$\widehat{\vartheta} = \arg\min_{\vartheta \in \Theta} \mathcal{J}(\vartheta). \tag{2.8}$$

The vector $\widehat{\vartheta}$ thus obtained is called the *weighted least-squares estimator*. It is natural to assume that $w_i = \varrho(x_i)$, i.e., to assign small weights to the points x_i where the variance of disturbances is relatively high.

If $\vartheta = \mathbb{R}^m$, it is an easy exercise to show that the necessary condition for the optimality of $\widehat{\vartheta}$ reduces to the system of linear equations

$$M_N \widehat{\vartheta} = b_N, \tag{2.9}$$

where

$$M_N = F_N^{\mathrm{T}} V_N F_N, \quad b_N = F_N^{\mathrm{T}} V_N y, \quad V_N = \mathrm{Diag}[\varrho(x_1), \ldots, \varrho(x_N)]. \tag{2.10}$$

M_N is usually called the Fisher Information Matrix (FIM). Consequently, the weighted least-squares estimate is

$$\widehat{\vartheta} = M_N^{-1} b_N = (F_N^{\mathrm{T}} V_N F_N)^{-1} F_N^{\mathrm{T}} V_N y \tag{2.11}$$

provided that M_N is not singular.

The least-squares criterion is only one of possible choices, but benefits from its use, especially valuable properties of the obtained estimators (unbiasedness, i.e., $E(\widehat{\vartheta}) = \vartheta^\star$, and efficiency, i.e., minimal variance), provide persuasive arguments to use it.

Also recall other important properties of the weighted least-squares estimator (Rafajłowicz, 2006):

- If the disturbances satisfy the assumption (2.4) and the weights in (2.7) are chosen such that $w_i = \varrho(x_i)/\sigma^2$, then the covariance matrix of the estimator $\widehat{\vartheta}$ is
$$\mathrm{Cov}(\widehat{\vartheta}) = E\left[(\widehat{\vartheta} - E\widehat{\vartheta})(\widehat{\vartheta} - E\widehat{\vartheta})^{\mathrm{T}}\right] = \sigma^2 M_N^{-1}. \tag{2.12}$$
- If the disturbances ε_i are Gaussian with zero mean and variance σ^2, then the estimator $\widehat{\vartheta}$ is Gaussian with mean $\vartheta^\star$ and covariance matrix $\sigma^2(F_N^{\mathrm{T}} V_N F_N)^{-1}$.
- The estimator of the regression function at point x is given by $\widehat{\eta}(x) = \widehat{\vartheta}^{\mathrm{T}} f(x)$. It is unbiased and
$$\mathrm{Var}(\widehat{\eta}(x)) = f(x)^{\mathrm{T}}\mathrm{Cov}(\widehat{\vartheta}) f(x) = \sigma^2 f(x)^{\mathrm{T}} M_N^{-1} f(x). \tag{2.13}$$

Remark 1. The maximum likelihood method is a popular choice for many practitioners (Seber and Wild, 1989; Bates and Watts, 1988). When we assume that the i-th observation is characterized by a probability density function $\mathrm{p}(y_i|x_i,\vartheta)$ and that observations are independent of one another, the maximum likelihood estimator has the form (Seber and Wild, 1989)
$$\widehat{\vartheta} = \arg\max_{\vartheta\in\Theta} \prod_{i=1}^{n} \mathrm{p}(y_i|x_i,\vartheta). \tag{2.14}$$

As indicated in (Fedorov and Hackl, 1997), the results obtained in experimental design theory when using least-squares estimators remain valid for the maximum likelihood estimators, since both the approaches lead to the same structure of the information matrix M_N. Some additional comments concerning benefits and drawbacks of using the least-squares criterion with further references can be found in the comprehensive monographs (Seber and Wild, 1989; Bates and Watts, 1988; Söderström and Stoica, 1988).

For any estimation method, it is obvious, that the applied observation strategy (here it is understood by the sequence $\{x_i\}$ which we informally call the simplified design of experiment and denote by ξ_N) affects the "accuracy" of the obtained estimates $\widehat{\vartheta}$ (since for different sets ξ_N we obtain different covariance matrices $\mathrm{Cov}(\widehat{\vartheta})$). This allows us to formulate the problem of observation scheduling so as to obtain "best" estimates of $\vartheta^\star$. To this end, it is necessary to construct some quality measure of observation schedules based on the accuracy of the parameter estimates obtained from the observations. In the classical theory of optimum experimental design such a measure is usually constructed based on the notion of the Fisher Information Matrix (FIM) (Fedorov and Hackl, 1997; Walter and Pronzato, 1997) which is defined as follows:
$$M_\vartheta = E_{y|\vartheta}\left\{\left[\frac{\partial \log \mathrm{p}(y|\vartheta)}{\partial\vartheta}\right]^{\mathrm{T}} \frac{\partial \log \mathrm{p}(y|\vartheta)}{\partial\vartheta}\right\}. \tag{2.15}$$

This is because, up to a constant, its inverse constitutes a lower bound of the covariance matrix for the estimates $\mathrm{Cov}(\widehat{\vartheta})$. This fact is expressed by the so-called

Cramér-Rao inequality which holds for any unbiased estimator $\widehat{\vartheta}$ (Goodwin and Payne, 1977; Söderström and Stoica, 1988):

$$\mathrm{Cov}(\widehat{\vartheta}) \succeq M_N^{-1}. \tag{2.16}$$

In the linear case considered here, the inequality in (2.16) becomes an equality, cf. (2.12). Note that the FIM (and its inverse) can be easily computed even in the situations when the exact dispersion matrix of a given estimator is very difficult to obtain. It is important to note that inequality (2.16) should be interpreted in terms of the Löwner ordering of symmetric matrices, i.e., $A \succeq B$ means that $A - B \succeq 0$ (i.e., $A - B$ must be non-negative definite). Direct comparison of covariance matrices or information matrices obtained for two different settings ξ_N^1 and ξ_N^2 can be difficult and inconvenient, since usually such matrices are not comparable in the sense of Löwner ordering ('$\succeq$' imposes solely partial ordering of information matrices). Thus, in order to compare different FIM's, it is indispensable to introduce the appropriate scalar performance index.

Before we present a family of such performance indices known in the literature, we have to formally define the notion of the design of an experiment and get the coresponding structure of the FIM.

2.1.2. Definitions of the experimental design

Definition 2.1. The table

$$\xi_N \stackrel{\text{def}}{=} \begin{Bmatrix} x_1, & x_2, & \dots, & x_n \\ p_1, & p_2, & \dots, & p_n \end{Bmatrix}, \tag{2.17}$$

where $p_i > 0$, $p_i = r_i/N$, $\sum_{i=1}^n p_i = 1$, $x_i \neq x_j$ for $i \neq j$, $\sum_{i=1}^n r_i = N$, $r_i > 0$ being the number of replicated observations at the location x_i, is denoted by ξ_N and called the *normalized N-observation exact design* of the experiment. The x_i's now are said to be the design or *support points*. Each p_i is called a *weight* and it can be interpreted as a proportion of the whole experimental effort which should be assigned to the corresponding support point x_i.

Observe that $p_i = r_i/N$ are non-negative rational numbers. If we relax this condition and allow p_i's to be any non-negative reals which sum up to unity, then we get an extremely useful generalization of the design, namely the so-called approximate discrete design of experiment:

Definition 2.2. The *approximate discrete* (or, simply, *approximate*) *design* of the experiment ξ concentrated at n support points is represented by the table

$$\xi = \begin{Bmatrix} x_1, & x_2, & \dots, & x_n \\ p_1, & p_2, & \dots, & p_n \end{Bmatrix}, \tag{2.18}$$

where the x_i's are elements of the design space X and the corresponding weights satisfy the conditions

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1. \tag{2.19}$$

We will denote by supp(ξ) the set of the support points $\{x_1, \ldots, x_n\}$.

Conditions (2.19) allow us to treat ξ as a probability distribution on the set $\{x_1, \ldots, x_n\}$, i.e., we may interpret ξ as a discrete random variable which takes values in the set $\{x_1, \ldots, x_n\}$ and whose probability mass function is defined by the corresponding weights $\{p_1, \ldots, p_n\}$.

The interpretation of the admissible designs as discrete probability distributions on finite subsets of X ameliorates to a great extent the tractability of the optimum experimental design problem. Nevertheless, there still remain many technicalities which make the resulting calculus rather cumbersome. These difficulties can be easily overcome by further widening the class of admissible designs to all probability measures ξ over the design space X.

Definition 2.3. The *approximate continuous* (or, simply, *continuous*) *design* is any measure ξ defined on the Borel sigma-field $\mathcal{B}(X)$ on X for which

$$\xi(X) = \int_X \xi(\mathrm{d}x) = 1. \tag{2.20}$$

The set of all continuous designs will be denoted by $\Xi(X)$.

Clearly, the integration with respect to ξ in (2.20) is to be understood in the Lebesgue-Stieltjes sense (Carter and van Brunt, 2000).

Note that a continuous design ξ can be non-zero on subsets of X consisting of single points, i.e., a particular point $x \in X$ may have a non-zero probability mass attached to it. More information on this generalization of the design and properties of continuous designs is introduced in Chapter 3.

2.1.3. Properties of Fisher Information Matrices

Based on (2.12) and (2.16), we can write down an expression defining the structure of the information matrix corresponding to a particular exact design ξ_N. If we use such a design and estimate parameters using the weighted least-squares method, we have $\mathrm{Cov}(\widehat{\vartheta}) = M_N^{-1}(\xi_N)$, where

$$M_N(\xi_N) = \sigma^{-2} N \sum_{i=1}^{n} \varrho(x_i) p_i f(x_i) f^{\mathrm{T}}(x_i). \tag{2.21}$$

In practice, it is customary to operate on the so-called normalized FIM

$$M(\xi_N) = \frac{\sigma^2}{N} M_N(\xi_N) = \sum_{i=1}^{n} \varrho(x_i) p_i f(x_i) f^{\mathrm{T}}(x_i) \tag{2.22}$$

instead of $M_N(\xi_N)$. We can then generalize the notion of the normalized information matrix to the case when we identify continuous designs as probability measures on X. Thus,

$$M(\xi) = \int_X \varrho(x) f(x) f^{\mathrm{T}}(x)\, \xi(\mathrm{d}x). \tag{2.23}$$

Information matrices for continuous designs possess the following properties which make the further optimization easier (Rafajłowicz, 2006):

- For each $\xi \in \Xi(X)$ the FIM $M(\xi)$ is symmetric and nonnegative definite, and has dimension $m \times m$, where $m = \dim(f(x))$.

- The information matrix $M(\xi)$ corresponding to a design ξ whose number of support points is less then m is singular.

- For any approximate discrete design, the information matrix is additive, i.e., it is the sum of information matrices that correspond to the individual observations:
$$M(\xi) = \sum_{i=1}^{n} p_i M(x_i), \tag{2.24}$$
where $M(x) = \varrho(x) f(x) f^{\mathrm{T}}(x)$. This important property, which holds also in nonlinear models, has significant meaning in practice, due to resulting computational savings.

- The set $\Xi(X)$ is convex. Thus, a convex combination of two designs $\xi^1 \in \Xi(X)$ and $\xi^2 \in \Xi(X)$ for $0 \le \alpha \le 1$ can be easily computed: $\xi = (1-\alpha)\xi^1 + \alpha\xi^2 \in \Xi(X)$.
In particular, for
$$\xi^1 = \begin{Bmatrix} x_1^1, & x_2^1, & \dots, & x_n^1 \\ p_1^1, & p_2^1, & \dots, & p_n^1 \end{Bmatrix} \tag{2.25}$$
and ξ^2 concentrated at a single point x^2,
$$\xi^2 = \begin{Bmatrix} x^2 \\ 1 \end{Bmatrix}, \tag{2.26}$$
we have
$$\begin{aligned} \xi &= (1-\alpha)\xi^1 + \alpha\xi^2 \\ &= \begin{Bmatrix} x_1^1, & x_2^1, & \dots, & x_n^1, & x^2 \\ p_1^1(1-\alpha), & p_2^1(1-\alpha), & \dots, & p_n^1(1-\alpha), & \alpha \end{Bmatrix} \end{aligned} \tag{2.27}$$
if $x^2 \notin \operatorname{supp}(\xi^1)$, and
$$\begin{aligned} \xi &= (1-\alpha)\xi^1 + \alpha\xi^2 \\ &= \begin{Bmatrix} x_1^1, & \dots, & x_{i-1}^1, & x_i^1, & & \\ p_1^1(1-\alpha), & \dots, & p_{i-1}^1(1-\alpha), & p_i(1-\alpha)+\alpha, & & \\ & & & x_{i+1}^1, & \dots, & x_n^1 \\ & & & p_{i+1}^1(1-\alpha), & \dots, & p_n^1(1-\alpha) \end{Bmatrix} \end{aligned} \tag{2.28}$$
if x^2 coincides with x_i^1 for some $i \in \{1, \dots, n\}$.

- Denote by $\mathcal{M}(X)$ the set of all information matrices ($\mathcal{M}(X) = \{M(\xi) : \xi \in \Xi(X)\}$). If $M_1, M_2 \in \mathcal{M}(X)$ then also $(\alpha M_1 + (1-\alpha)M_2) \in \mathcal{M}$ for each $0 \leq \alpha \leq 1$. Indeed, M_1 and M_2 belong to $\mathcal{M}(X)$ only if there exist designs $\xi_1, \xi_2 \in \Xi$ such that $M_1 = M(\xi^1)$ and $M_2 = M(\xi^2)$. Thus,

$$\alpha M_1 + (1-\alpha)M_2 = M[\alpha\xi^1 + (1-\alpha)\xi^2] \tag{2.29}$$

  Since the set of all continuous designs is convex, $(\alpha\xi^1 + (1-\alpha)\xi^2) \in \Xi(X)$, which implies $M[\alpha\xi^1 + (1-\alpha)\xi^2] \in \mathcal{M}(X)$, i.e., the set $\mathcal{M}(X)$ is indeed convex.

- If the design region X is compact and the functions $f(\cdot)$ and $\varrho(\cdot)$ are continuous on X, then the set $\mathcal{M}(X)$ is also compact (Pázman, 1986; Uciński, 2005). This property is of paramount importance, since it assures the existence of solutions to the problem of optimum experimental design.

- Let the assumptions of the last property remain valid. Then for each design $\xi \in \Xi(X)$ there exists an equivalent approximate discrete design $\xi' \in \Xi(X)$ such that $M(\xi) = M(\xi')$ and ξ' consists of at most $m(m+1)/2+1$ distinct support points. Moreover, if $M(\xi)$ lies at the boundary of the set $\mathcal{M}(X)$, then an equivalent approximate discrete design ξ' can be found containing at most $m(m+1)/2$ points. This property results from Carathéodory's Theorem (Silvey, 1980; Rockafellar, 1997).

2.2. Optimizing the design quality

In this section we present a few optimality criteria which are commonly used in optimum experimental design to compare different designs.

2.2.1. Comparison of different designs

From the Gauss-Markov theorem (Caines, 1988) it is well known that the least squares estimator is the best one in the whole class of linear unbiased estimators in the sense of the Löwner ordering of the covariance matrices. Moreover, the unbiased linear predictor $\widehat{y}(x) = \widehat{\vartheta}^{\mathrm{T}} f(x)$ of the regression function $\vartheta^{\mathrm{T}} f(x)$ has minimal variance in the same class.

The Gauss-Markov theorem is important, since we know that we cannot improve the accuracy in the estimates by simply changing the method of data processing. Thus we can:

- decrease the variance of disturbances $\sigma^2(x)\varrho^{-1}(x)$, which requires either more precise measurement devices or multiple replication of the same observation and averaging,

- increase the number of measurements N, but usually it also increases the total cost and time of performing of the experiment,

- increase the experimental region X, but it is sometimes just impossible or connected with additional costs.

If none of the above manners is applicable or we need a further increase in the accuracy of the estimates, all possibilities reduce to an optimal selection of the experimental design ξ_N provided that N and X are fixed and the variance of observations cannot be decreased.

If the assumptions concerning the estimation method, model and noise structure made in Sec. 2.1.1 still hold, then it can be shown that the likelihood (confidence) ellipsoid for parameter estimates of $\vartheta^\star$ has the structure

$$(\vartheta - \widehat{\vartheta})^{\mathrm{T}} M_N(\xi_N)(\vartheta - \widehat{\vartheta}) \leq c, \tag{2.30}$$

where c denotes some positive *constant* dependent on the number of observations N, the number of parameters m, the estimated variance of disturbances and a confidence level $0 < \beta < 1$. The centre of such an ellipsoid is situated at the point $\widehat{\vartheta}$ and covers the region of the estimated parameters values with probability β. Changes in the design ξ may influence the orientation of the axes and their lengths. It is obvious that the "smaller" the ellipsoid (in an adapted sense), the more accurate the parameter estimates and, consequently, the design for which the ellipsoid is "smaller" is better. Below, we present some frequently used design criteria motivated by various properties of the concentration ellipsoid, where $\xi^1 \preceq \xi^2$ denotes that a design ξ^1 is not better than ξ^2 in the sense of the criterion at hand. Below, we follow the notation of (Rafajłowicz, 2006).

- **Evaluation based on the volume of the likelihood ellipsoid**.
  The volume of the ellipsoid (2.30) is proportional to $\left[\det(M^{-1}(\xi))\right]^{1/2}$. Thus, based on the determinant of the covariance matrix, we regard the design ξ^1 as no better than ξ^2 when

$$\xi^1 \overset{D}{\preceq} \xi^2 \quad \text{iff} \quad \det[M^{-1}(\xi^1)] \geq \det[M^{-1}(\xi^2)], \tag{2.31}$$

  where the letter 'D' is used as an abbreviation for *determinant*.

- **Evaluation based on the average axis length of the likelihood ellipsoid**.
  The average axis length of the ellipsoid (2.30) is proportional to the trace of $M^{-1}(\xi)$. Thus,

$$\xi^1 \overset{A}{\preceq} \xi^2 \quad \text{iff} \quad \operatorname{trace}[M^{-1}(\xi^1)] \geq \operatorname{trace}[M^{-1}(\xi^2)], \tag{2.32}$$

  where the letter 'A' derives from the word *average*.

- **Evaluation based on the average p-th power of the axis length**.
  The method constitutes a generalization of the previous one:

$$\xi^1 \overset{L_p}{\preceq} \xi^2 \quad \text{iff} \quad \left(\operatorname{trace}[M^{-p}(\xi^1)]\right)^{1/p} \geq \left(\operatorname{trace}[M^{-p}(\xi^2)]\right)^{1/p}. \tag{2.33}$$

- **Evaluation based on the linearly weighted mean**.
  The method is actually another generalization of the one based on the average length of the axes of the concentration ellipsoid:

$$\xi^1 \overset{L}{\preceq} \xi^2 \quad \text{iff} \quad \operatorname{trace}[AM^{-1}(\xi^1)] \geq \operatorname{trace}[AM^{-1}(\xi^2)], \tag{2.34}$$

  where a non-negative definite matrix A is chosen *a priori*.

- **Evaluation based on the maximum axis length of the likelihood ellipsoid**.
  The length of the i-th axis is equal to

$$2\sqrt{\lambda_i(M^{-1}(\xi))}, \tag{2.35}$$

  where $\lambda_i(\,\cdot\,)$, $i = 1, \ldots, m$ denote the eigenvalues of its matrix argument. Thus,

$$\xi^1 \overset{E}{\preceq} \xi^2 \quad \text{iff} \quad \max_i \lambda_i(M^{-1}(\xi^1)) \geq \max_i \lambda_i(M^{-1}(\xi^2)), \tag{2.36}$$

  where the letter 'E' is used as an abbreviation for *eigenvalue*.

The foregoing designs are aimed at improving the accuracy of the parameter estimates. Alternatively, one could be interested in design comparison with respect to the quality of estimating the value of the regression function. For unbiased estimators of y the accuracy of the estimates at a point x is measured using the variance of prediction $\operatorname{Var}(\widehat{y}(x))$. This variance also depends on the employed observation schedule (i.e., the design ξ). Indeed, setting $\varphi(x, \xi) = \operatorname{Var}(\widehat{y}(x))$, we have

$$\varphi(x, \xi) = f(x)^{\mathrm{T}} M^{-1}(\xi) f(x). \tag{2.37}$$

Most popular functionals constructed based on the function $\varphi(x, \xi)$ are listed below:

- **Extrapolation at point x_0**.
  If a point x_0 is given *a priori* and we are interested in a more accurate estimation of the regression function value, then we can consider the following choice:

$$\xi^1 \preceq \xi^2 \quad \text{iff} \quad \varphi(x_0, \xi^1) \geq \varphi(x_0, \xi^2). \tag{2.38}$$

- **Evaluation based on the average variance of the output**.
  Since the variance function is non-negative, its integration gives information about its average values. Thus, we could simply set

$$\xi^1 \overset{Q}{\preceq} \xi^2 \quad \text{iff} \quad \int_X \varphi(x_0, \xi^1)\,\mathrm{d}x \geq \int_X \varphi(x_0, \xi^2)\,\mathrm{d}x. \tag{2.39}$$

  Note that this criterion is equivalent to (2.34) when we choose the matrix A such that $A = \int_X f(x) f(x)^{\mathrm{T}} \mathrm{d}x$.

- **Evaluation based on the maximum variance of the output**.
  Here we base on the assumption that the quality of the design is measured by the variance of the output at a point where $\varphi(x,\xi)$ attains its maximum value. Thus,

$$\xi^1 \overset{G}{\preceq} \xi^2 \quad \text{iff} \quad \max_{x\in X}\varphi(x,\xi^1) \geq \max_{x\in X}\varphi(x,\xi^2). \tag{2.40}$$

  This idea of design ordering is strictly connected with (2.31) (cf. the next section)

Remark 2. The designs for which $\xi^1 \preceq \xi^2$ and simultaneously $\xi^1 \succeq \xi^2$ are said to be equivalent. For example, for the ordering relation $\overset{D}{\preceq}$ this means the equal determinants of their covariance (and also information) matrices for both the designs. Generally, we say that two designs are equivalent in the problem of parameter estimation of the regression function, and denote this fact by $\xi^1 \equiv \xi^2$, when

$$M(\xi^1) = M(\xi^2). \tag{2.41}$$

2.2.2. Generic properties of design criteria

The idea of design ordering presented in the last section can be easily generalized. Bearing in mind considerations presented at the end of Section 2.1.1, let Ψ denote a scalar function defined on $m \times m$ information matrices and $\mathrm{NND}(m)$ denote the set of all symmetric, non-negative definite $m \times m$ matrices. The general conditions which the function $\Psi : \mathrm{NND}(m) \to \mathbb{R}$ must usually satisfy in order to be used as an optimality criterion are as follows (Fedorov and Hackl, 1997):

- **Homogeneity**. For any $\alpha > 0$ and matrix $A \in \mathrm{NND}(m)$ we have $\Psi(\alpha A) = s(\alpha)\Psi(A)$ where s denotes some real-valued, non-negative and nondecreasing function. As a result, for a matrix $M_N(\xi_N)$ of the form (2.21), we have

$$\Psi(M_N) = \Psi(N\sigma^{-2}M) = s(N\sigma^{-2})\Psi(M), \tag{2.42}$$

  i.e., maximization of $\Psi(M_N)$ amounts to that of $\Psi(M)$.

- **Monotonicity**. For any matrices $A, B \in \mathrm{NND}(m)$ satisfying $A \succeq B$ we have $\Psi(A) \geq \Psi(B)$.

- **Concavity**. For any $0 \leq \alpha \leq 1$ and $A, B \in \mathrm{NND}(m)$, we have

$$\Psi[(1-\alpha)A + \alpha B] \geq (1-\alpha)\Psi(A) + \alpha\Psi(B). \tag{2.43}$$

- **Differentiability**. For any nonsingular matrix $A \in \mathrm{NND}(m)$ there exists an $m \times m$ matrix $\overset{\circ}{\Psi}(A)$ such that

$$\overset{\circ}{\Psi}(A) = \left[\frac{\partial\Psi(A)}{\partial a_{ij}}\right]. \tag{2.44}$$

The introduction of an optimality criterion Ψ makes it possible to formulate the problem of optimum experimental design for regression as an optimization one: A design $\xi^\star$ is called Ψ-optimum iff

$$\xi^\star = \arg \max_{\xi \in \Xi(X)} \Psi(M(\xi)). \tag{2.45}$$

The most popular choices of the corresponding criteria are listed below:

- D-optimality criterion
$$\Psi(M) = \ln \det(M), \tag{2.46}$$
- A-optimality criterion
$$\Psi(M) = -\mathrm{trace}(M^{-1}), \tag{2.47}$$
- E-optimality criterion
$$\Psi(M) = \lambda_{\min}(M), \tag{2.48}$$
where $\lambda_{\min}(\,\cdot\,)$ denotes the minimal eigenvalue of its matrix argument,
- The sensitivity criterion
$$\Psi(M) = \mathrm{trace}(M). \tag{2.49}$$
In contrast to the predecessors, this criterion does not possess a statistical interpretation in the context of the likelihood ellipsoid. The name is motivated by the fact that its maximization yields the increased sensitivity of the outputs with respect to parameter changes.

Another class of criteria is formed by functions which refer to the variance of the prediction of the system output (Dette and O'Brien, 1999). The most important one is defined as follows:

- G-optimality criterion
$$\Psi(M) = -\max_{x \in X} \left\{ f(x)^{\mathrm{T}} M^{-1} f(x) \right\}. \tag{2.50}$$
We thus minimize the maximal variance of the prediction for the response. This criterion is closely connected to the D-optimality one, as shown in the sequel.

Now we present the fundamental result of the experimental design theory for regression, i.e., Kiefer and Wolfowitz's Equivalence Theorem. It reveals the equivalence between D- and G-optimum designs and provides conditions which allow us to check whether or not a design $\xi^\star$ is D- or G-optimal. Moreover, it plays a vital role in the construction of numerical algorithms used for finding optimum designs.

Theorem 2.1 (Kiefer-Wolfowitz)**.** *The following design problems are equivalent: Find $\xi^\star$ so that*

1. $\xi^\star = \arg \max_{\xi \in \Xi(X)} \det(M)$,
2. $\xi^\star = \arg \min_{\xi \in \Xi(X)} \max_{x \in X} \psi(x, \xi)$,
3. $\max_{x \in X} \psi(x, \xi^\star) = m$,
   where $\psi(x, \xi) = f^{\mathrm{T}}(x) M^{-1}(\xi) f(x)$.

The relevant optimum design $\xi^\star$ exists and we can always find an equivalent design which is approximate discrete with at most $m(m+1)/2$ support points. Moreover, at all support points $x_i^\star$, $i = 1, \ldots, n$ of a purely discrete D-optimum design $\xi^\star$ the function $\psi(x, \xi^\star)$ attains its maximum which equals the number of the estimated parameters, i.e., $\psi(x_i^\star, \xi^\star) = m$.

The proof of the above theorem, based on differential optimality conditions for convex functions can be found in numerous monographs (Fedorov, 1972; Fedorov and Hackl, 1997). There exists a generalized version of Theorem 2.1 which gives optimality conditions for a general optimality criterion Ψ satisfying the assumptions discussed at the beginning of this section (see, e.g., Theorem 2.3.2 from (Fedorov and Hackl, 1997)). A lot of modifications of the above mentioned generalization, established for particular criteria exist, all being also called the equivalence theorems.

Example 2.1. Using the Kiefer-Wolfowitz Theorem it is easy to show that the design of the form

$$\xi^\star = \begin{Bmatrix} -1, & 1 \\ 1/2, & 1/2 \end{Bmatrix} \tag{2.51}$$

is D-optimal for the linear model $\eta(x, \vartheta) = \vartheta_1 + \vartheta_2 x$ when the design region is $X = [-1, 1]$. Indeed, since $M(\xi^\star) = I_2$, the variance function has the form $\psi(x, \xi^\star) = 1 + x^2$ and attains its maximum value on X at the support points ± 1. Moreover, that maximum equals just the number of parameters $m = 2$.

2.3. Nonlinear models

A more complicated situation takes place when the considered system is not linear with respect to the parameters. Now we replace the linear response function $\vartheta^{\mathrm{T}} f(x)$ by a possibly nonlinear response $\eta(x, \vartheta)$, where $\eta : X \times \Theta \to \mathbb{R}$ and $\Theta \subset \mathbb{R}^m$ is the set of admissible parameters. On the analogy of (2.8), the least-squares estimates of model parameters are

$$\widehat{\vartheta} = \arg \min_{\vartheta \in \Theta} \frac{1}{2} \sum_{i=1}^{N} w_i \left(y_i - \eta(x_i, \vartheta) \right)^2 . \tag{2.52}$$

It is generally not possible to find a closed-form expression for $\widehat{\vartheta}$ in most practical situations. Moreover, the estimator (2.52) is generally biased (i.e., $E\widehat{\vartheta} \neq \vartheta^\star$) and except for special cases, it is extremely difficult to find an analytical expression for the covariance matrix $\mathrm{Cov}(\widehat{\vartheta})$. To derive an approximation to the dispersion matrix, it is customary to linearize the system response in the vicinity of a prior estimate ϑ^0 of the unknown parameter vector ϑ (Walter and Pronzato, 1997). If ϑ^0 is close enough to the true value $\vartheta^\star$, then the system response can be approximated by expanding the function $\eta(x,\vartheta)$ in the Taylor series with respect to ϑ. Then, retaining only linear terms, we have

$$\eta(x,\vartheta) \cong \eta(x,\vartheta^0) + \left.\frac{\partial\eta(x,\vartheta)}{\partial\vartheta}\right|_{\vartheta=\vartheta^0} (\vartheta - \vartheta^0) \tag{2.53}$$

In this way we obtain a counterpart of the observation equation (2.2) for the nonlinear case:

$$y_i = \eta(x_i,\vartheta^0) + \left.\frac{\partial\eta(x_i,\vartheta)}{\partial\vartheta}\right|_{\vartheta=\vartheta^0} (\vartheta - \vartheta^0) + \varepsilon_i, \quad i = 1,\ldots,N. \tag{2.54}$$

For $w_i = \sigma^{-2}\varrho(x_i)$, the corresponding form of the Fisher Information Matrix being analogous to (2.22) is

$$M(\xi) = \sum_{i=1}^{n} \varrho(x_i) g(x_i) g(x_i)^{\mathrm{T}}, \tag{2.55}$$

where

$$g(x) = \left(\frac{\partial\eta(x,\vartheta)}{\partial\vartheta}\right)^{\mathrm{T}}_{\vartheta=\vartheta^0} \tag{2.56}$$

and is called the sensitivity vector. Its 'continuous' version corresponding to (2.23) is

$$M(\xi) = \int_X \varrho(x) g(x) g(x)^{\mathrm{T}}\, \xi(\mathrm{d}x). \tag{2.57}$$

It is clear that the information matrix (2.55) depends on the prior estimate ϑ^0 around which the model is linearized and is valid only when the approximation (2.53) is accurate enough. In this sense the results obtained from such an approach have only local character. However, it can be shown that, under rather mild assumptions, the estimator (2.52) is strongly consistent (Seber and Wild, 1989), i.e.,

$$a.s.\lim_{N\to\infty} \widehat{\vartheta} = \vartheta^\star, \tag{2.58}$$

where $a.s.\lim$ denotes the *almost sure* limit (convergence with probability one). Thus, if both $\eta(x,\vartheta)$ and $g(x)$ are continuous on $X\times\Theta$ and X, respectively, then all the results incorporated in the previous sections are directly applicable without any changes. In particular, we substitute $g(x)$ for $f(x)$, e.g., the G-optimality criterion takes the form

$$\Psi(M) = -\max_{x\in X} \left\{ g(x)^{\mathrm{T}} M^{-1} g(x) \right\}. \tag{2.59}$$

2.4. Numerical approximation of optimum designs

First algorithms for finding numerical approximations to D-optimum designs were presented at the beginning of the 1970s (Fedorov, 1972; Wynn, 1970). In the literature they are collectively called the Wynn-Fedorov algorithm. Below we present the idea and basic scheme of such a procedure in the case of non-linear models.

2.4.1. Sequential construction of D-optimum designs

Assume that we have a design ξ^0 such that $\det(M(\xi^0)) > 0$. Our goal is to find a design ξ^1 which will increase the value of the D-optimality criterion as much as possible. Because of this, consider a set of designs $(1-\alpha)\xi^0 + \alpha\xi_{new}$ for $0 \leq \alpha \leq 1$ and ξ_{new} being some design based on which we wish to improve (increase) the value of $\det(M(\xi))$. Then we can define the univariate function

$$\omega(\alpha) = \Psi\left[(1-\alpha)\xi^0 + \alpha\xi_{new}\right]. \tag{2.60}$$

Its right-hand derivative at 0 is given by

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\Psi\left[(1-\alpha)\xi^0 + \alpha\xi_{new}\right]\Big|_{\alpha=0+} = \operatorname{trace}[M^{-1}(\xi^0)M(\xi_{new})] - m \tag{2.61}$$

and indicates the ascent of the FIM determinant in the neighborhood of ξ^0 when we move into the direction of ξ_{new}. The search for a convenient candidate ξ_{new} is a highly complex optimization problem *per se*, but we may restrict the class of such designs to the ones concentrated at single points, i.e., the designs of the form $\xi_{new} = \left\{ {}^{x_{new}}_{1} \right\}$, where $x_{new} \in X$. Then we have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\Psi\left[(1-\alpha)\xi^0 + \alpha\xi_{new}\right]\Big|_{\alpha=0+} = \varrho(x_{new})g^{\mathrm{T}}(x_{new})M^{-1}(\xi^1)g(x_{new}) - m, \tag{2.62}$$

where $g(x)$ is defined by (2.56).

Just as in the steepest-ascent method (Bertsekas, 1999), now the goal is to find a point $x_{new} \in X$ for which the value of $\varrho(x_{new})g^{\mathrm{T}}(x_{new})M^{-1}(\xi^1)g(x_{new})$ will be as large as possible. Consequently, such a choice of x_{new} leads to a **locally** highest improvement in the value of the FIM determinant. This idea forms a basis for the following algorithm.

2.4.2. Wynn-Fedorov algorithm for D-optimum designs

Algorithm 2.1 (Wynn-Fedorov algorithm for D-optimum designs)**.**

Step 1: Choose an initial non-singular design ξ^0 (thus it must include at least m support points) and some positive tolerance $\epsilon \ll 1$. Set $k = 1$.

Step 2: Find

$$x^k = \arg\max_{x \in X} \psi(x, \xi^k), \tag{2.63}$$

where

$$\psi(x,\xi) = \varrho(x) g^{\mathrm{T}}(x) M^{-1}(\xi) g(x). \tag{2.64}$$

Step 3: If $\psi(x_k, \xi^k) \leq m(1+\epsilon)$, then $\xi^\star = \xi^k$, STOP. Otherwise, go to Step 4.

Step 4: Choose an appropriate value of α^k with $0 \leq \alpha^k \leq 1$ and compute the convex combination of designs:

$$\xi^{k+1} = (1-\alpha^k)\xi^k + \alpha^k \delta(x^k) \tag{2.65}$$

where $\delta(x^k)$ is the design concentrated at x^k, i.e., $\delta(x^k) = \left\{ \begin{smallmatrix} x^k \\ 1 \end{smallmatrix} \right\}$.
Set $k \leftarrow k+1$ and go to Step 2.

The choice of the sequence α^k is not unambiguous. It can be shown that in the considered case of single-response systems, the optimum value (assuring the highest improvement in the D-criterion value) along the line joining ξ^k with $\delta(x^k)$ is given by (Fedorov, 1972; Fedorov and Hackl, 1997)

$$\alpha^k = \frac{\psi(x^k, \xi^k) - m}{(\psi(x^k, \xi^k) - 1)m}. \tag{2.66}$$

When a generalization to multi-response systems is considered, the choice of an optimum α^k value necessitates a line-search procedure (e.g., the simple golden search (Patan, 2004)). Another common choice is simply setting (Wu and Wynn, 1978)

$$\alpha^k = \frac{1}{k+1}. \tag{2.67}$$

In general, α^k can be chosen using any sequence satisfying the conditions:

$$\lim_{k\to\infty} \alpha^k = 0, \quad \sum_{k=0}^{\infty} \alpha^k = \infty, \quad \sum_{k=0}^{\infty} (\alpha^k)^2 < \infty. \tag{2.68}$$

Step 2 constitutes the most critical part of the algorithm, since it involves a search for a global optimum of a multi-modal function. There exist many modifications of this basic scheme which aim at reducing the total computational cost. An interesting approach, called the selective random search, was presented in (Rafajłowicz, 1998; Rafajłowicz, 2006). The idea relies on the observation that, given ξ^k, the sensitivity function $\psi(x, \xi^k)$ after a suitable normalization can be treated as a probability distribution function on the design region X. Thus, instead of looking for $x^k = \arg\max_{x \in X} \psi(x, \xi^k)$, we generate its approximation as a random number sampled from this distribution and satisfying $\psi(x, \xi^k) > m$.

The presented algorithm belongs to the group of first-order methods, exploiting information only about the gradient of the target function. In practice, support points close to the optimum ones are selected relatively fast (during a few up to a dozen or so iterations) but a precise weight determination takes much more time. Thus, it is worth to speed up this process by additionally optimizing weights during each iteration of the Wynn-Fedorov procedure. A helpful procedure is described below.

2.4.3. Sequential algorithm for designs on a finite set $\boldsymbol{X}$

There exists a modified version of the Wynn-Fedorov scheme which is suited for constructing D-optimum designs on a finite set of distinct support points given *a priori*. Here, however, we describe a much faster multiplicative algorithm (Titterington, 1976; Silvey *et al.*, 1978; Pázman, 1986; Pukelsheim and Torsney, 1991).

Assume that we have n fixed points $x_1, \ldots, x_n$. Then, the design $\xi \in \Xi(X)$, where $X = \{x_1, \ldots, x_n\}$, is determined by the set of weights $p_1, \ldots, p_n$ such that $\sum_{i=1}^{n} p_i = 1$. Then consider the following mapping T which transforms the design ξ into $\xi' = T(\xi)$ by modifying only the weights in such a way that

$$p_i' = p_i \frac{\psi(x_i, \xi)}{m}, \quad i = 1, \ldots, n. \tag{2.69}$$

Observe that the Kiefer-Wolfowitz theorem implies that for the optimum design $\psi(x_i, \xi^\star)/m = 1, \quad i = 1, \ldots, n$. Therefore the weight vector of the optimum design $\xi^\star$ constitutes a fixed point of the mapping T. We can prove the global convergence of the following algorithm (Pázman, 1986; Uciński, 2005):

Algorithm 2.2 (Weight optimization algorithm for D-optimum designs)**.**

Step 1: Choose an initial non-singular design ξ^0 such that $p_i^0 > 0$, $i = 1, \ldots, n$. Select $\epsilon \ll 1$, some positive tolerance. Set $k = 0$.

Step 2: If $\psi(x_i, \xi^k) \leq m(1 + \epsilon)$, $i = 1, \ldots, n$, then set $\xi^\star = \xi^k$ and STOP.

Step 3: Calculate the design ξ^{k+1} by modifying the weights of ξ^k as follows:

$$p_i^{k+1} = p_i^k \frac{\psi(x_i, \xi^k)}{m}, \quad i = 1, \ldots, n. \tag{2.70}$$

Set $k \leftarrow k + 1$ and go back to Step 2.

This algorithm can be further improved by on-line elimination of points which fail a special test based on ellipsoidal trimming introduced to identify support points having no chance to be included in an optimal design (Pronzato, 2003).

2.5. Summary

The material contained in this chapter is aimed at introducing the reader to some elements of modern optimum experimental design. It will be extensively used in the next parts of this dissertation, since the methodologies of characterizing and numerically determining T-optimum designs were introduced in the mid-1970s by drawing an analogy with those for smooth criteria, such as the D-optimality one. This constitutes perhaps a main reason for the reported problems with the adopted Wynn-Fedorov scheme, cf. (Fedorov and Hackl, 1997), which have been tolerated since the seminal paper by (Atkinson and Fedorov, 1975a). In the sequel, we shall show that the T-optimum criterion requires a special treatment and conceive new efficient algorithms for its maximization.

Chapter 3

DESIGN OF EXPERIMENTS TO DISCRIMINATE BETWEEN RIVAL MODELS

A major problem in the optimum experimental design theory is connected with structure identification (Burnham and Anderson, 2002; Söderström and Stoica, 1988). It is a common situation in modelling real processes that we obtain several plausible models and we wish to discriminate between them. An optimally designed experiment allows us for certainty maximization of such discrimination. It also minimizes experimental effort, which is a valuable property when the measurements are expensive or difficult to collect.

3.1. Problem set-up

In the following, we consider a basic scheme when two concurrent models of a static multi-response process are of interest. This corresponds to the situation in which observations y_{ij} are given by

$$y_{ij} = \eta(x_i) + \varepsilon_{ij}, \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where the function $\eta : \mathbb{R}^s \to \mathbb{R}^d$ constitutes the *true* model of the process. In this description x_i stands for a setting of the vector $x \in X \subset \mathbb{R}^s$ of independent variables (the so-called *explanatory* or *regressor* variables), e.g., a time instant and/or a spatial coordinate of a measurement, $x_i \neq x_\kappa$ whenever $i \neq \kappa$, X being some known compact set. The terms ε_{ij} represent random fluctuations and possibly errors of measurement resulting from inaccuracies in the measuring devices. The errors ε_{ij} are sampled from a distribution satisfying

$$E\big[\varepsilon_{ij}\big] = 0, \quad E\big[\varepsilon_{ij}\varepsilon_{\kappa\ell}^{\mathrm{T}}\big] = \begin{cases} \sigma^2 I_d & \text{if } i = \kappa \text{ and } j = \ell, \\ 0_d & \text{otherwise,} \end{cases} \tag{3.2}$$

where I_d is the $d \times d$ identity matrix, 0_d is the $(d \times d)$-dimensional matrix of zeros and σ^2 is a constant positive variance. Without loss of generality, in the theoretical developments we can take σ^2 to be unity.

The additional index j is necessary if the observations are to be repeated several times for some settings x_i, as in practice repeated experimental runs typically lead to different observed responses. Here the number of replications for a given setting x_i is denoted by r_i and we have $\sum_{i=1}^{n} r_i = N$.

The x_i's are called control variables because they can be chosen by the experimenter. They define process conditions and may vary from observation to observation (in (3.1) we thus have n different settings denoted by $x_1, \ldots, x_n$).

Our basic assumption is that the model response $\eta(x)$ coincides with either $\eta_1(x, \vartheta_1)$ or $\eta_2(x, \vartheta_2)$, where $\mathbb{R}^d$-valued, possibly nonlinear, functions η_1 and η_2 are given *a priori*, $\vartheta_1 \in \Theta_1 \subset \mathbb{R}^{m_1}$ and $\vartheta_2 \in \Theta_2 \subset \mathbb{R}^{m_2}$ being constant parameters which are fixed but unknown to the experimenter (Θ_1 and Θ_2 denote some known compact sets).

The purpose of the experiment is to determine which of the models η_1 and η_2 is true. This goal can be realized using the approach based on the notion of T-optimality (on which we are focus in the main part of the dissertation).

As is well known (Fedorov and Hackl, 1997; Walter and Pronzato, 1997), cf. also (2.52), the least-squares estimates of the parameters ϑ_ℓ are respectively given by

$$\widehat{\vartheta}_{\ell N} = \arg \min_{\vartheta_\ell \in \Theta_\ell} \sum_{i=1}^{n} p_i \left\| y_i - \eta_\ell(x_i, \vartheta_\ell) \right\|^2, \quad \ell = 1, 2, \tag{3.3}$$

where

$$y_i = \frac{1}{r_i} \sum_{j=1}^{r_i} y_{ij}, \quad p_i = \frac{r_i}{N}. \tag{3.4}$$

There is no loss of generality in assuming that the first model is true, i.e., $\eta(x) \equiv \eta_1(x, \vartheta_1^\star)$, where a fixed value of $\vartheta_1^\star$ is regarded as known prior to the experiment (this value could be obtained based on some preliminary experiment or some nominal value could be used if available). To make the discrimination between the models η_1 and η_2 as accurate as possible then amounts to selection of x_i's and p_i's so as to maximize the so-called non-centrality parameter defined as follows (Atkinson and Fedorov, 1975a):

$$\Delta_1^0(\xi_N) = \min_{\vartheta_2 \in \Theta_2} \mathcal{J}^0(\xi_N, \vartheta_2), \tag{3.5}$$

where

$$\mathcal{J}^0(\xi_N, \vartheta_2) = \sum_{i=1}^{n} p_i \left\| \eta(x_i) - \eta_2(x_i, \vartheta_2) \right\|^2 \tag{3.6}$$

and ξ_N stands for the normalized design of experiment in the form (2.17).

The criterion defined in such a manner is called the T-optimum design one to emphasize a close connection with hypothesis testing (the letter 'T' comes from 'testing'), since discrimination between the competing models $\mathcal{M}_1$ and $\mathcal{M}_2$ can be viewed as testing the following simple non-parametrical hypothesis:

$$\mathcal{H}_0 : \quad \eta(x_i) = \eta_1(x_i, \vartheta_1), \quad i = 1, \ldots, N, \tag{3.7}$$

against the alternative

$$\mathcal{H}_1 : \quad \eta(x_i) = \eta_2(x_i, \vartheta_2), \quad i = 1, \ldots, N. \tag{3.8}$$

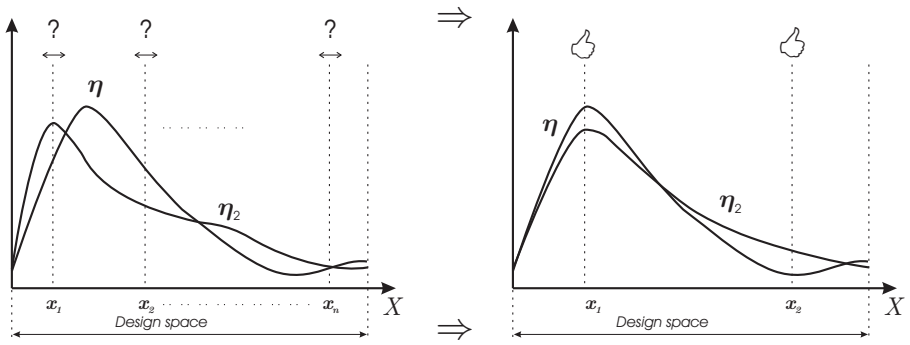


Fig. 3.1. Idea of T-optimum design.

Remark 3. In practise, for many complex industrial processes, to allow for repetitions of observations at particular support points is very inconvenient, expensive or even impossible to realize. Nevertheless, the outlined approach is not useless, as the weights p_i could be interpreted as precisions of measuring devices applied to take measurements at particular points. A more systematic approach which can be used in such situations and which exploits the notion of replication-free designs, will be considered later.

Motivations behind the criterion (3.5) are rather intuitively clear (see Fig. 3.1), as it constitutes a measure of discrepancy between the responses of both the models: a good design for discriminating between the models will then provide a large lack of fit in terms of the sum of squares for the second model. What is more, they are also confirmed by theory, since for normally distributed errors ε_{ij} and large N, the noncentrality parameter $\Delta_1^0(\xi_N)$ is proportional (in mean) to the logarithm of the ratio of the likelihoods associated with both the models. We shall demonstrate this property later in this chapter.

Clearly, the solutions (designs) depend on the choice of the true model, as well as on the true values of its parameters (i.e., $\vartheta_1^\star$). This implies that a construction of *a-priori* solutions is impossible and we are faced with a problem which is analogous to the search for the so-called *local designs* while trying to optimally estimate unknown parameters of a nonlinear regression function. We wish to choose a design which is optimal for some particular value of ϑ_1 in the hope that it is not too bad whatever the true $\vartheta_1^\star$ happens to be. The dependency of the optimal solution on the model's parameters is an unappealing characteristic of nonlinear experimental designs (Walter and Pronzato, 1997; Atkinson and Donev, 1992). This predicament can be partially circumvented by relying on a nominal value of ϑ_1, the results of a preliminary experiment or a sequential design which consists in multiple alternation of experimentation and estimation steps.

Remark 4. The experimenter can possess some prior information about the distribution of the model (models) parameters or even probabilities of each candidate model to be the true one. That valuable information can be used to construct

designs which are 'robust' to parametric uncertainty using a Bayesian approach (Atkinson, 1992; Ponce de Leon and Atkinson, 1991).

Note that, similarly to (2.17), the design ξ_N in (3.5) defines a discrete probability distribution on its support points $x_1, \ldots, x_n$, since the corresponding weights fulfil the conditions

$$p_i \geq 0, \quad i = 1, \ldots, n, \tag{3.9}$$

$$\sum_{i=1}^{n} p_i = 1. \tag{3.10}$$

Thus the p_i's are rational numbers, since both the r_i's and N are integers. This discrete nature of N-observation exact designs causes serious difficulties, as the resultant numerical analysis problem is not amenable to solution by standard optimization techniques, particularly when N is large. A commonly used device for this problem is to extend the definition of the design. When N is large, the p_i's can be considered as continuous in the interval $[0, 1]$ (and not necessarily integer multiples of $1/N$). Therefore, we might think of the designs as all discrete probability distributions on X for which the number of support points n is not fixed and constitutes an additional parameter to be determined. This makes the problem much more tractable since we have the opportunity of exploiting calculus techniques in solving it. Having the number of support points n and associated values of $x_i^\star$, together with the proportions of measurements assigned to those support points $p_i^\star$ where the maximum of the T-optimality criterion occurs, we can argue that the maximum of the same function over the integer multiples of $1/N$ occurs at adjacents to $p_1^\star, \ldots, p_n^\star$. The above reinterpretation of the admissible designs as discrete probability distributions on finite subsets of X ameliorates to a significant measure the tractability of the optimum experimental design problem. Nevertheless, there still remain many technicalities which make the resulting calculus rather cumbersome. Paradoxically, those difficulties can be easily overcome by widening the class of admissible designs a bit further, i.e., to all probability measures ξ defined on the Borel sigma-field on X, cf. Section 2.1.2, where the same generalization was made for other design criteria. By definition they must satisfy

$$\int_X \xi(\mathrm{d}x) = 1, \tag{3.11}$$

where the integration is to be understood in the Lebesgue-Stieltjes sense. The set of all such measures will be denoted by $\Xi(X)$. This relaxation of the original optimization problem usually leads only to approximate solutions, but these are generally accepted in the practice of optimum experimental design. What is more, this idea leads to some elegant and useful theory.

The corresponding 'relaxed' admissible designs are called *continuous* or *approximate* designs and they constitute the foundation of the modern theory of optimal experiments (Silvey, 1980; Cox and Reid, 2000; Atkinson and Donev, 1992; Pázman, 1986; Fedorov and Hackl, 1997) which originates from seminal works by (Kiefer and Wolfowitz, 1959).

Accordingly, as a continuous generalization of the noncentrality parameter, we consider the quantity

$$\Delta_1(\xi) = \min_{\vartheta_2 \in \Theta_2} \mathcal{J}(\xi, \vartheta_2), \tag{3.12}$$

where

$$\mathcal{J}(\xi, \vartheta_2) = \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x). \tag{3.13}$$

A design satisfying

$$\xi^\star = \arg \max_{\xi \in \Xi(X)} \Delta_1(\xi) \tag{3.14}$$

will then be called the *locally T-optimum design*.

Remark 5. Continuous designs with support points $x_1, x_2, \ldots, x_n$ supply information about a relative distribution of measurement frequency at $x_1, x_2, \ldots, x_n$. In order to use such designs in practice, it is necessary to transform these frequencies into numbers of experiments which are to be repeated at each support point. Such a procedure can be implemented as follows:

- Choose a total number of observations $N > 0$.
- Calculate the numbers $r_i^{'} = Np_i$ for each $i = 1, 2, \ldots, n$.
- Perform the rounding as follows:
  - compute the number of repetitions at the i-th point $r_i^{''} = \lfloor r_i^{'} \rfloor, \quad i = 1, 2, \ldots, n$
    ($\lfloor u \rfloor$ denotes the greatest integer no greater than u),
  - compute $N_r = N - \sum_{i=1}^{n} r_i^{''}$,
  - assign N_r remaining measurements to points $x_1, x_2, \ldots, x_n$ (e.g., using random sampling with replacement).

Using the above procedure, we obtain the numbers of observations which should be acquired at each point x_i. The order of acquiring information is not unambiguous. Depending on the situation at hand it can be realized

- randomly, which, according to Fisher's suggestions, helps to justify the assumption of normal errors (Atkinson and Bailey, 2001),
- systematically, which can be applied when the randomization cannot be used or when its use gives additional benefits such as a decrease in the cost or simplification of computations.

A detailed analysis of the accuracy of the rounding procedure can be found in (Pukelsheim, 1993).

3.2. Properties of the T-optimality criterion

In what follows, we provide several important interpretations and characterizations of the T-optimality criterion. They turn out to be of paramount importance for numerical algorithms which are going to be outlined in the forthcoming chapters.

3.2.1. A statistical justification for the T-optimality criterion

For simplicity, assume that only one response can be observed, i.e., $d = 1$, and then consider the design

$$\xi_N = \begin{Bmatrix} x_1, & \ldots, & x_N \\ 1/N, & \ldots, & 1/N \end{Bmatrix} \tag{3.15}$$

for some finite N and $x_i \in X$, $i = 1, \ldots, N$, and purely random Gaussian measurement errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. When model $\mathcal{M}_1$ is fitted to the data $y_1, \ldots, y_N$, the corresponding likelihood function becomes

$$\mathcal{L}(\vartheta_1|\xi_N, \mathcal{M}_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(y_i - \eta_1(x_i, \vartheta_1))^2}{\sigma^2}\right). \tag{3.16}$$

In the same situation, but assuming that model $\mathcal{M}_2$ is true, we would obtain the likelihood

$$\mathcal{L}(\vartheta_2|\xi_N, \mathcal{M}_2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(y_i - \eta_2(x_i, \vartheta_2))^2}{\sigma^2}\right). \tag{3.17}$$

Consequently, the ratio of $\mathcal{L}(\vartheta_1|\xi_N, \mathcal{M}_1)$ to $\mathcal{L}(\vartheta_2|\xi_N, \mathcal{M}_2)$ is

$$\begin{aligned} L &= \frac{\mathcal{L}(\vartheta_1|\xi_N, \mathcal{M}_1)}{\mathcal{L}(\vartheta_2|\xi_N, \mathcal{M}_2)} \\ &= \exp\left\{\frac{1}{2\sigma^2}\left[\sum_{i=1}^{N}(y_i - \eta_2(x_i, \vartheta_2))^2 - \sum_{i=1}^{N}(y_i - \eta_1(x_i, \vartheta_1))^2\right]\right\}. \end{aligned} \tag{3.18}$$

Fixing ϑ_ℓ at the value of

$$\widehat{\vartheta}_\ell^N = \arg\min_{\vartheta_\ell \in \Theta_\ell} \sum_{i=1}^{N}(y_i - \eta_\ell(x_i, \vartheta_\ell))^2, \quad \ell = 1, 2, \tag{3.19}$$

we see that

$$\begin{aligned} E\left[(y_i - \eta_\ell(x_i, \widehat{\vartheta}_\ell^N))^2\right] &= E\left[(\eta(x_i) - \eta_\ell(x_i, \widehat{\vartheta}_\ell^N) + y_i - \eta(x_i))^2\right] \\ &= (\eta(x_i) - \eta_\ell(x_i, \widehat{\vartheta}_\ell^N))^2 + E\left[(y_i - \eta(x_i))^2\right]. \end{aligned} \tag{3.20}$$

Consequently,

$$\begin{aligned} 2\sigma^2 E[\log L] &= E\left[\sum_{i=1}^{N}(y_i - \eta_2(x_i, \widehat{\vartheta}_2^N))^2 - \sum_{i=1}^{N}(y_i - \eta_1(x_i, \widehat{\vartheta}_1^N))^2\right] \\ &= \sum_{i=1}^{N}(\eta(x_i) - \eta_2(x_i, \widehat{\vartheta}_2^N))^2 + \sum_{i=1}^{N} E\left[(y_i - \eta(x_i))^2\right] \\ &\quad - \sum_{i=1}^{N}(\eta(x_i) - \eta_1(x_i, \widehat{\vartheta}_1^N))^2 - \sum_{i=1}^{N} E\left[(y_i - \eta(x_i))^2\right] \\ &= N\Delta_1(\xi_N) - \sum_{i=1}^{N}(\eta(x_i) - \eta_1(x_i, \widehat{\vartheta}_1^N))^2, \end{aligned} \tag{3.21}$$

where

$$\Delta_1(\xi_N) = \frac{1}{N}\sum_{i=1}^{N}(\eta(x_i) - \eta_2(x_i, \widehat{\vartheta}_2^N))^2. \tag{3.22}$$

If $\mathcal{M}_1$ is true, then $\eta(x)$ coincides with $\eta_1(x, \vartheta_1^0)$ for some $\vartheta_1^0 \in \Theta_1$. Under some mild and natural regularity conditions, cf. Theorem 1.1 in (Ermakov and Zhigljavsky, 1987, p. 197), the estimator $\widehat{\vartheta}_1^N$ is then strongly consistent, i.e., $\widehat{\vartheta}_1^N \xrightarrow[N\to\infty]{} \vartheta_1^0$ almost surely. Moreover, the sequence

$$\sum_{i=1}^{N}(\eta(x_i) - \eta_1(x_i, \widehat{\vartheta}_1^N))^2 \tag{3.23}$$

converges in probability to a constant as $N \to \infty$ and thus it is stochastically bounded. Therefore, asymptotically,

$$E(\log L) \sim \frac{N}{2\sigma^2}\Delta_1(\xi_N). \tag{3.24}$$

For large N (in principle, this is the case considered in this dissertation), maximization of $\Delta_1(\xi)$ thus leads to maximization in the average sense of the logarithm of the likelihood ratio $\mathcal{L}(\vartheta_1|\cdot, \mathcal{M}_1)/\mathcal{L}(\vartheta_2|\cdot, \mathcal{M}_2)$. On the other hand, for linear models the logarithm of this likelihood ratio provides a basis for justification that $N\Delta_1(\xi)/\sigma^2$ is the non-centrality parameter of the χ^2 distribution of the residual sum of squares for $\mathcal{M}_2$ and that the T-optimum design maximizing $\Delta_1(\xi)$ provides the most powerful F-test for lack of fit of $\mathcal{M}_2$ when $\mathcal{M}_1$ is true. When the models are non-linear in the parameters, the exact F-test is replaced by asymptotic results, but we still design to maximize $\Delta_1(\xi)$ (Atkinson and Donev, 1992).

3.2.2. Weak continuity of $\Delta_1(\cdot)$

Further, we present our main theoretical results for which we assume the following:

(A1) X and Θ_2 are compact sets.

(A2) $\eta(\,\cdot\,)$ is a continuous function on X.

(A3) $\eta_2(\,\cdot\,,\,\cdot\,)$ is a continuous function on $X \times \Theta_2$.

Here and subsequently, we will use the symbol $\Xi(X)$ to denote the set of all probability measures on $\mathcal{B}(X)$, the σ-algebra of Borel subsets of X.

To shorten notation, we let $e(x, \vartheta_2)$ stand for $\|\eta(x) - \eta_2(x, \vartheta_2)\|^2$ and write $\widehat{\Theta}_2(\xi)$ instead of $\operatorname{Arg}\min_{\vartheta_2 \in \Theta_2} \int_X e(x, \vartheta_2)\, \xi(\mathrm{d}x)$. It is easily seen that $e(\cdot, \cdot)$ is continuous.

We begin with a definition and an auxiliary result which are not commonly known. We include them for the convenience of the reader, thus making our exposition self-contained.

Definition 3.1. Let $\mathcal{B}(X)$ denote the σ-algebra of Borel subsets of a complete and separable metric space X. A sequence of probability measures $\{\mu_k\}$, defined on $\mathcal{B}(X)$, is said to *converge weakly* to a probability measure μ, on $\mathcal{B}(X)$, if for every continuous and bounded function $g : X \to \mathbb{R}$ we have

$$\int_X g(x)\, \mu_k(\mathrm{d}x) \xrightarrow[k\to\infty]{} \int_X g(x)\, \mu(\mathrm{d}x). \tag{3.25}$$

In this case we write '$\mu_k \to \mu$ weakly'. Note that the limit μ is unique.

Lemma 3.1. *(Ermakov and Zhigljavsky, 1987, Lemma 1.4, p. 91)*
If $X \subset \mathbb{R}^s$ is compact, then every sequence of probability measures on $\mathcal{B}(X)$ has a subsequence that converges weakly to a probability measure on $\mathcal{B}(X)$.

The above result, being an immediate conclusion drawn from Prohorov's Theorem (Pollard, 2002) means the so-called sequentially weak compactness of $\Xi(X)$.

The main focus of our study in this section will be the 'continuity' property of $\Delta_1(\,\cdot\,)$. To this end, we need the following technical result:

Lemma 3.2. *Let the Assumptions (A1)–(A3) hold. Suppose that a sequence $\{\xi^k\}$ of $\Xi(X)$ converges weakly to an element $\widehat{\xi} \in \Xi(X)$ and a sequence $\{\vartheta_2^k\}$ of $\mathbb{R}^m$ converges to an element $\widehat{\vartheta}_2 \in \mathbb{R}^m$. Then*

$$\lim_{k\to\infty} \int_X e(x, \vartheta_2^k)\, \xi^k(\mathrm{d}x) = \int_X e(x, \widehat{\vartheta}_2)\, \widehat{\xi}(\mathrm{d}x). \tag{3.26}$$

Proof. From Assumptions (A1)–(A3) it follows that the function $e(\cdot, \cdot)$ is uniformly continuous on $X \times \Theta_2$. Therefore, given $\epsilon > 0$, we must have for some index $\bar{k} \geq 1$

$$e(x, \widehat{\vartheta}_2) - \epsilon < e(x, \vartheta_2^k) < e(x, \widehat{\vartheta}_2) + \epsilon, \quad \forall x \in X,\, \forall k \geq \bar{k}. \tag{3.27}$$

Integrating over X with respect to ξ^k yields

$$\int_X e(x, \widehat{\vartheta}_2)\, \xi^k(\mathrm{d}x) - \epsilon < \int_X e(x, \vartheta_2^k)\, \xi^k(\mathrm{d}x) < \int_X e(x, \widehat{\vartheta}_2)\, \xi^k(\mathrm{d}x) + \epsilon, \quad \forall k \geq \bar{k}. \tag{3.28}$$

Letting $k \to \infty$, we see that

$$\begin{aligned}\int_X e(x,\widehat{\vartheta}_2)\,\widehat{\xi}(\mathrm{d}x) - \epsilon &\le \liminf_{k\to\infty} \int_X e(x,\vartheta_2^k)\,\xi^k(\mathrm{d}x) \\ &\le \limsup_{k\to\infty} \int_X e(x,\vartheta_2^k)\,\xi^k(\mathrm{d}x) \\ &\le \int_X e(x,\widehat{\vartheta}_2)\,\widehat{\xi}(\mathrm{d}x) + \epsilon.\end{aligned} \tag{3.29}$$

Since ϵ was selected arbitrarily, we get

$$\int_X e(x,\vartheta_2^k)\,\xi^k(\mathrm{d}x) \xrightarrow[k\to\infty]{} \int_X e(x,\widehat{\vartheta}_2)\,\widehat{\xi}(\mathrm{d}x). \tag{3.30}$$

□

We can now formulate our main result.

Theorem 3.1. *Under Assumptions (A1)–(A3), the T-optimality criterion* $\Delta_1(\,\cdot\,)$ *is weakly continuous, i.e., if* $\xi^k \to \widehat{\xi}$ *weakly in* $\Xi(X)$*, then* $\Delta_1(\xi^k) \to \Delta_1(\widehat{\xi})$.

Proof. We first show that $\Delta_1(\,\cdot\,)$ is weakly lower semicontinuous, i.e., that for any sequence $\{\xi^k\}$ in $\Xi(X)$ such that $\xi^k \to \widehat{\xi}$ weakly, as $k \to \infty$,

$$\liminf_{k\to\infty} \Delta_1(\xi^k) \ge \Delta_1(\widehat{\xi}). \tag{3.31}$$

Note that the limit on the left-hand side of (3.31) is finite owing to the boundedness of $\Delta_1(\,\cdot\,)$, which results in turn from the continuity of $e(\cdot,\cdot)$ on the compact set $X \times \Theta_2$.

For each $k = 1, 2, \ldots$ define ϑ_2^k such that

$$\Delta(\xi^k) = \int_X e(x,\vartheta_2^k)\,\xi^k(\mathrm{d}x). \tag{3.32}$$

Let $\{\xi^k\}_{\mathcal{K}^0}$, $\mathcal{K}^0 \subset \mathbb{N}$ be a subsequence of $\{\xi^k\}$ for which we have

$$\liminf_{k\to\infty} \Delta_1(\xi^k) = \lim_{\substack{k\to\infty \\ k\in\mathcal{K}^0}} \Delta_1(\xi^k). \tag{3.33}$$

The compactness of Θ_2 implies that we can select a subsequence $\{\xi^k\}_{\mathcal{K}}$, $\mathcal{K} \subset \mathcal{K}^0$, such that $\{\vartheta_2^k\}_{\mathcal{K}} \to \vartheta_2^\star$ for some $\vartheta_2^\star \in \Theta_2$. Consequently,

$$\begin{aligned}\Delta_1(\widehat{\xi}) &\le \int_X e(x,\vartheta_2^\star)\,\widehat{\xi}(\mathrm{d}x) && \text{by definition} \\ &= \lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \int_X e(x,\vartheta_2^k)\,\xi^k(\mathrm{d}x) && \text{by Lemma 3.2} \\ &= \liminf_{k\to\infty} \Delta_1(\xi^k) && \text{by (3.33)},\end{aligned} \tag{3.34}$$

which is the desired conclusion.

Now we shall prove the weak upper semicontinuity of $\Delta_1(\,\cdot\,)$, which means that for any sequence $\{\xi^k\}$ in $\Xi(X)$ such that $\xi^k \to \widehat{\xi}$ weakly, as $k \to \infty$,

$$\limsup_{k\to\infty} \Delta_1(\xi^k) \le \Delta_1(\widehat{\xi}). \tag{3.35}$$

The existence of a finite limit on the left-hand side of (3.35) results from the boundedness of $\Delta_1(\,\cdot\,)$ on $\Xi(X)$.

For the sake of contradiction, suppose that there is a sequence $\{\xi^k\}$ in $\Xi(X)$ such that $\xi^k \to \widehat{\xi}$ weakly and

$$\lim_{k\to\infty} \Delta_1(\xi^k) = \Delta_1(\widehat{\xi}) + \delta, \tag{3.36}$$

for some $\delta > 0$. Choosing $\widehat{\vartheta}_2 \in \widehat{\Theta}_2(\widehat{\xi})$, we have

$$\Delta_1(\widehat{\xi}) = \int_X e(x, \widehat{\vartheta}_2)\, \widehat{\xi}(\mathrm{d}x). \tag{3.37}$$

Since $\{\xi^k\}$ converges weakly to $\widehat{\xi}$, we see that

$$\lim_{k\to\infty} \int_X e(x, \widehat{\vartheta}_2)\, \xi^k(\mathrm{d}x) = \int_X e(x, \widehat{\vartheta}_2)\, \widehat{\xi}(\mathrm{d}x). \tag{3.38}$$

Hence there exists an index $\bar{k} \ge 1$ such that

$$\begin{aligned} \int_X e(x, \widehat{\vartheta}_2)\, \xi^k(\mathrm{d}x) &< \int_X e(x, \widehat{\vartheta}_2)\, \widehat{\xi}(\mathrm{d}x) + \frac{\delta}{2} \\ &= \Delta_1(\widehat{\xi}) + \frac{\delta}{2} < \Delta_1(\xi^k), \quad \forall k \ge \bar{k}, \end{aligned} \tag{3.39}$$

which contradicts the definition of $\Delta_1(\xi^k)$.

By (3.31) and (3.35) it is obvious that $\Delta_1(\,\cdot\,)$ must be weakly continuous. □

3.2.3. Weak outer semicontinuity of $\widehat{\Theta}_2(\,\cdot\,)$

This section discusses some properties of the set-valued function

$$\widehat{\Theta}_2(\xi) = \operatorname{Arg}\min_{\vartheta_2 \in \Theta_2} \int_X e(x, \vartheta_2)\, \xi(\mathrm{d}x). \tag{3.40}$$

The set $\widehat{\Theta}_2(\xi)$ is sometimes called the 'answering set': the elements ϑ_2 in $\widehat{\Theta}_2(\xi)$ 'answer ξ' (Wong, 1992). The relevant properties will turn out to be of great importance while examining the convergence of the relaxation algorithm outlined in Chapter 4.

Theorem 3.2. *Under Assumptions (A1)–(A3), for each $\xi \in \Xi(X)$ the set $\widehat{\Theta}_2(\xi)$ is compact.*

Proof. The result follows easily from the continuity of the function

$$g(\vartheta_2) = \int_X e(x, \vartheta_2)\, \xi(\mathrm{d}x) \tag{3.41}$$

defined on Θ_2 and the immediate observation that

$$\widehat{\Theta}_2(\xi) = \operatorname{Arg} \min_{\vartheta_2 \in \Theta_2} g(\vartheta_2). \tag{3.42}$$

□

In the following definition, we adopt the common terminology regarding set-valued functions (Polak, 1997, p. 676).

Definition 3.2. Let $\{\xi^k\}$ be a sequence in $\Xi(X)$.

(a) The point $\widehat{\vartheta}_2 \in \Theta_2$ is said to be a *limit point* of $\{\widehat{\Theta}_2(\xi^k)\}$ if

$$\inf_{\vartheta_2 \in \widehat{\Theta}_2(\xi^k)} \|\vartheta_2 - \widehat{\vartheta}_2\| \xrightarrow[k\to\infty]{} 0. \tag{3.43}$$

(b) The point $\widehat{\vartheta}_2 \in \Theta_2$ is a *cluster point* of $\{\widehat{\Theta}_2(\xi^k)\}$ if it is a limit point of a subsequence of of $\{\widehat{\Theta}_2(\xi^k)\}$.

(c) We denote the set of all cluster points of $\{\widehat{\Theta}_2(\xi^k)\}$ by $\operatorname{Lim\,sup}_{k\to\infty} \widehat{\Theta}_2(\xi^k)$ and call it the *outer limit* of $\{\widehat{\Theta}_2(\xi^k)\}$.

The following result plays a central role while studying the convergence of numerical algorithms outlined in this dissertation.

Theorem 3.3. *Let Assumptions (A1)–(A3) hold. The mapping* $\widehat{\Theta}_2(\,\cdot\,)$ *is* ***sequentially weakly outer semicontinuous****, i.e., given* $\widehat{\xi} \in \Xi(X)$ *and any sequence* $\{\xi^k\}$ *such that* $\xi^k \to \widehat{\xi}$ *weakly, as* $k \to \infty$*, we have*

$$\operatorname{Lim\,sup}_{k\to\infty} \widehat{\Theta}_2(\xi^k) \subset \widehat{\Theta}_2(\widehat{\xi}). \tag{3.44}$$

Proof. Suppose the assertion of the theorem is false, i.e., there exists a point $\widehat{\vartheta}_2 \in \widehat{\Theta}_2(\widehat{\xi})$ and a sequence $\xi^k \to \widehat{\xi}$ weakly, as $k \to \infty$, such that, for some $\vartheta_2^k \in \widehat{\Theta}_2^k(\xi^k)$, we have $\vartheta_2^k \to \widehat{\vartheta}_2 \notin \widehat{\Theta}_2(\widehat{\xi})$, as $k \to \infty$. But then Lemma 3.2 implies

$$\Delta_1(\xi^k) = \int_X e(x, \vartheta_2^k)\, \xi^k(\mathrm{d}x) \xrightarrow[k\to\infty]{} \int_X e(x, \widehat{\vartheta}_2)\, \widehat{\xi}^k(\mathrm{d}x) > \Delta_1(\widehat{\xi}), \tag{3.45}$$

the last inequality resulting from the non-optimality of $\widehat{\vartheta}_2$. This, however, contradicts the weak continuity of $\Delta_1(\,\cdot\,)$. □

3.2.4. Concavity of $\Delta_1(\,\cdot\,)$

For any probability measures ξ_1 and ξ_2 on $\mathcal{B}(X)$ and any α with $0 \le \alpha \le 1$, we have that $\xi = (1-\alpha)\xi_1 + \alpha\xi_2$ is a probability measure on $\mathcal{B}(X)$ too (for it is immediate that $\int_X \xi(\mathrm{d}x) = (1-\alpha)\int_X \xi_1(\mathrm{d}x) + \alpha \int_X \xi_2(\mathrm{d}x) = 1$). This means that $\Xi(X)$ is convex. We call this ξ the *convex combination* of ξ_1 and ξ_2.

The following result will be crucial for characterizing solutions of the T-optimum design problem:

Theorem 3.4. *Under Assumptions (A1)–(A3), the function $\Delta_1(\,\cdot\,)$ is concave, i.e., for all $\xi_1, \xi_2 \in \Xi(X)$, and α with $0 \le \alpha \le 1$, we have*

$$\Delta_1((1-\alpha)\xi_1 + \alpha\xi_2) \ge (1-\alpha)\Delta_1(\xi_1) + \alpha\Delta_1(\xi_2). \tag{3.46}$$

Proof. Let us fix some $\xi_1, \xi_2 \in \Xi(X), \alpha \in [0,1]$ and let $\xi = (1-\alpha)\xi_1 + \alpha\xi_2$. For every $\vartheta_2 \in \Theta_2$, we have

$$\int_X e(x,\vartheta_2)\,\xi(\mathrm{d}x) = (1-\alpha)\int_X e(x,\vartheta_2)\,\xi_1(\mathrm{d}x) + \alpha\int_X e(x,\vartheta_2)\,\xi_2(\mathrm{d}x). \tag{3.47}$$

Taking the minimum over all $\vartheta_2 \in \Theta_2$, we conclude that

$$\begin{aligned}\Delta_1(\xi) &= \min_{\vartheta_2\in\Theta_2} \int_X e(x,\vartheta_2)\,\xi(\mathrm{d}x) \\ &\ge (1-\alpha)\min_{\vartheta_2\in\Theta_2}\int_X e(x,\vartheta_2)\,\xi_1(\mathrm{d}x) + \alpha \min_{\vartheta_2\in\Theta_2}\int_X e(x,\vartheta_2)\,\xi_2(\mathrm{d}x) \\ &= (1-\alpha)\Delta_1(\xi_1) + \alpha\Delta_1(\xi_2),\end{aligned} \tag{3.48}$$

which is our claim. □

3.2.5. Existence of a T-optimum design

Clearly, before proceeding to maximize the T-optimality criterion, it is useful to know that there exists at least one global maximum of $\Delta_1(\,\cdot\,)$ over $\Xi(X)$. This is guaranteed by the following result:

Theorem 3.5. *Let Assumptions (A1)–(A3) hold. Then there is an element $\xi^\star \in \Xi(X)$ such that*

$$\Delta_1(\xi^\star) = \sup_{\xi\in\Xi(X)} \Delta_1(\xi). \tag{3.49}$$

Proof. The continuity of $e(\cdot,\cdot)$ on $X \times \Theta_2$ yields

$$c = \sup_{\xi\in\Xi(X)} \Delta_1(\xi) < \infty. \tag{3.50}$$

Let $\{\xi^k\}$ be a maximizing sequence of elements in (3.50), i.e.,

$$\Delta_1(\xi^k) \uparrow c \quad \text{as} \quad k \to \infty. \tag{3.51}$$

Lemma 3.1 then implies that there exists a subsequence $\{\xi^k\}_{\mathcal{K}}, \mathcal{K} \subset \mathbb{N}$ of $\{\xi^k\}$ and a probability measure $\widehat{\xi} \in \Xi(X)$ such that $\{\xi^k\}_{\mathcal{K}} \to \widehat{\xi}$ weakly. From Theorem 3.1 it follows that $\{\Delta_1(\xi^k)\}_{\mathcal{K}} \to \Delta_1(\widehat{\xi})$. This, taken in conjunction with (3.50), gives $c = \Delta_1(\widehat{\xi})$, which is the desired conclusion. □

Note that the element $\xi^\star$ maximizing $\Delta_1(\,\cdot\,)$ may not be unique. What is more, Theorem 3.8 will show that the set of all maximizers of $\Delta_1(\,\cdot\,)$ is convex.

3.2.6. Directional derivative of $\Delta_1(\,\cdot\,)$

Modern optimum experimental design theory makes extensive use of the powerful tool which is the directional derivative of the optimality criterion. Similarly, we define the one-sided directional derivative of $\Delta_1(\,\cdot\,)$ at a point $\xi \in \Xi(X)$ in the direction $\mu \in \Xi(X)$ by

$$\mathrm{d}\Delta_1(\xi;\mu) = \lim_{\alpha \downarrow 0} \frac{\Delta_1((1-\alpha)\xi + \alpha\mu) - \Delta_1(\xi)}{\alpha}. \tag{3.52}$$

The existence of the relevant limit is guaranteed by the following result:

Theorem 3.6. *Under Assumptions (A1)–(A3), the directional derivative* $\mathrm{d}\Delta_1(\xi;\mu)$ *exists for all* $\xi, \mu \in \Xi(X)$ *and is given by*

$$\mathrm{d}\Delta_1(\xi;\mu) = \min_{\vartheta_2 \in \widehat{\Theta}_2(\xi)} \int_X \|\eta(x) - \eta_2(x,\vartheta_2)\|^2 \,\mu(\mathrm{d}x) - \Delta_1(\xi) \tag{3.53}$$

Proof. Fix $\xi, \mu \in \Xi(X)$ and define the function $f : [0,1] \to \mathbb{R}$ by

$$f(\alpha) = \min_{\vartheta_2 \in \Theta_2} \phi(\alpha, \vartheta_2), \tag{3.54}$$

where

$$\phi(\alpha,\vartheta_2) = (1-\alpha) \int_X e(x,\vartheta_2)\,\xi(\mathrm{d}x) + \alpha \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x). \tag{3.55}$$

Then $\mathrm{d}\Delta_1(\xi;\mu)$ coincides with the right-hand derivative $D_+f(\alpha)$ of $f(\,\cdot\,)$ at $\alpha = 0$, i.e.,

$$\mathrm{d}\Delta_1(\xi;\mu) = D_+f(0) = \lim_{\alpha \downarrow 0} \frac{f(\alpha) - f(0)}{\alpha}. \tag{3.56}$$

But $D_+f(0)$ can be easily determined based on Danskin's Theorem (Bertsekas, 1999, Prop. B.25, p. 717):

$$D_+f(0) = \min_{\vartheta_2 \in \widetilde{\Theta}_2(0)} \left[\frac{\partial}{\partial\alpha}\phi(\alpha,\vartheta_2)\right]_{\alpha=0^+}, \tag{3.57}$$

where

$$\widetilde{\Theta}_2(\alpha) = \left\{\vartheta_2 \in \Theta_2 : \phi(\alpha,\vartheta_2) = \min_{\vartheta_2 \in \Theta_2} \phi(\alpha,\vartheta_2)\right\}. \tag{3.58}$$

We check out at once that

$$\frac{\partial}{\partial\alpha}\phi(\alpha,\vartheta_2) = \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) - \int_X e(x,\vartheta_2)\,\xi(\mathrm{d}x). \tag{3.59}$$

Observing that $\widetilde{\Theta}_2(0) = \widehat{\Theta}_2(\xi)$, we finally get

$$\begin{aligned} \mathrm{d}\Delta_1(\xi;\mu) &= \min_{\vartheta_2\in\widehat{\Theta}_2(\xi)} \left\{ \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) - \int_X e(x,\vartheta_2)\,\xi(\mathrm{d}x) \right\} \\ &= \min_{\vartheta_2\in\widehat{\Theta}_2(\xi)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) - \Delta_1(\xi). \end{aligned} \tag{3.60}$$

□

If $\widehat{\Theta}_2(\xi)$ is a singleton, i.e., it consists of a unique point $\widehat{\vartheta}_2$, $\widehat{\Theta}_2(\xi) = \{\widehat{\vartheta}_2\}$, then

$$\mathrm{d}\Delta_1(\xi;\mu) = \int_X \|\eta(x) - \eta_2(x,\widehat{\vartheta}_2)\|^2\,\mu(\mathrm{d}x) - \Delta_1(\xi). \tag{3.61}$$

Furthermore, if μ is the unit-weight design concentrated at a single point $x \in X$, i.e., μ is the one-point (Dirac) measure $\delta(x)$, then (3.61) simplifies to

$$\mathrm{d}\Delta_1(\xi;\delta(x)) = \min_{\vartheta_2\in\widehat{\Theta}_2(\xi)} \|\eta(x) - \eta_2(x,\vartheta_2)\|^2 - \Delta_1(\xi). \tag{3.62}$$

Note that, in general, we have

$$\mathrm{d}\Delta_1(\xi;\mu) \neq \int_X \mathrm{d}\Delta_1(\xi;\delta(x))\,\mu(\mathrm{d}x), \tag{3.63}$$

which means that $\Delta_1(\,\cdot\,)$ does not meet the condition of the so-called linear differentiability, the property which substantially simplifies the analysis of smooth optimum design criteria (Fedorov and Hackl, 1997).

3.2.7. Equivalence theorem for the T-optimality criterion

Hardly anyone would doubt that optimality conditions are fundamental to the analysis of an optimization problem. In our design problem, the concavity of $\Delta_1(\,\cdot\,)$ over the convex set $\Xi(X)$ makes it possible to derive then a convenient test for the T-optimality of intuitively sensible designs.

In what follows, we find a characterization of $\xi^\star$ which has the property $\Delta_1(\xi^\star) = \max_{\xi\in\Xi(X)} \Delta_1(\xi)$.

Theorem 3.7 (Equivalence theorem for T-optimality designs)**.** *Let Assumptions (A1)–(A3) hold. Then, a necessary and sufficient condition for $\xi^\star$ to maximize $\Delta_1(\,\cdot\,)$ over $\Xi(X)$ is the existence of a probability measure $\zeta^\star$ on the σ-algebra of Borel subsets of the answering set $\widehat{\Theta}_2(\xi^\star)$ such that, for all $x \in X$,*

$$\int_{\widehat{\Theta}_2(\xi^\star)} \|\eta(x) - \eta_2(x,\vartheta_2)\|^2\,\zeta^\star(\mathrm{d}\vartheta_2) \leq \Delta(\xi^\star). \tag{3.64}$$

Proof. From the concavity of $\Delta_1(\,\cdot\,)$ over $\Xi(X)$ it follows that

$$\mathrm{d}\Delta_1(\xi^\star;\mu) \le 0, \quad \forall \mu \in \Xi(X) \tag{3.65}$$

constitutes a necessary and sufficient condition for $\xi^\star$ to maximize $\Delta_1(\,\cdot\,)$ over $\Xi(X)$. (The proof is standard, cf. e.g., (Uciński, 2005, Thm. B.25, p. 266)). By Theorem 3.6, this characterization becomes

$$\min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \le \Delta_1(\xi^\star), \quad \forall \mu \in \Xi(X), \tag{3.66}$$

or equivalently,

$$\max_{\mu \in \Xi(X)} \min_{\vartheta_2 \in \widehat{\theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \le \Delta_1(\xi^\star). \tag{3.67}$$

Also, setting $\Sigma(\xi^\star)$ as the set of all probability measures on the σ-algebra of Borel subsets of $\widehat{\Theta}_2(\xi^\star)$, and $\Sigma_0(\xi^\star)$ as its subset consisting of one-point Dirac measures, we get

$$\begin{aligned}
&\min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \\
&\qquad = \min_{\zeta \in \Sigma_0(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x)\,\zeta(\mathrm{d}\vartheta_2) \\
&\qquad \ge \min_{\zeta \in \Sigma(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x)\,\zeta(\mathrm{d}\vartheta_2) \\
&\qquad \ge \min_{\zeta \in \Sigma(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \left[\min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \right] \zeta(\mathrm{d}\vartheta_2) \\
&\qquad = \left[\min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \right] \left[\min_{\zeta \in \Sigma(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \zeta(\mathrm{d}\vartheta_2) \right] \\
&\qquad = \min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x),
\end{aligned} \tag{3.68}$$

which leads to

$$\min_{\vartheta_2 \in \widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) = \min_{\zeta \in \Sigma(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x)\,\zeta(\mathrm{d}\vartheta_2). \tag{3.69}$$

Similarly, we can show that

$$\max_{\mu \in \Xi(X)} \int_X \int_{\widehat{\Theta}_2(\xi^\star)} e(x,\vartheta_2)\,\zeta(\mathrm{d}\vartheta_2)\,\mu(\mathrm{d}x) = \max_{x \in X} \int_{\widehat{\Theta}_2(\xi^\star)} e(x,\vartheta_2)\,\zeta(\mathrm{d}\vartheta_2). \tag{3.70}$$

From (3.69) and (3.70) we have that

$$\begin{aligned}
&\max_{\mu\in\Xi(X)} \min_{\vartheta_2\in\widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x) \\
&\quad = \max_{\mu\in\Xi(X)} \min_{\zeta\in\Sigma(\xi^\star)} \int_{\widehat{\Theta}_2(\xi^\star)} \int_X e(x,\vartheta_2)\,\mu(\mathrm{d}x)\,\zeta(\mathrm{d}\vartheta_2) \\
&\quad = \min_{\zeta\in\Sigma(\xi^\star)} \max_{\mu\in\Xi(X)} \int_X \int_{\widehat{\Theta}_2(\xi^\star)} e(x,\vartheta_2)\,\zeta(\mathrm{d}\vartheta_2)\,\mu(\mathrm{d}x) \\
&\quad = \min_{\zeta\in\Sigma(\xi^\star)} \max_{x\in X} \int_{\widehat{\Theta}_2(\xi^\star)} e(x,\vartheta_2)\,\zeta(\mathrm{d}\vartheta_2),
\end{aligned} \tag{3.71}$$

where, additionally, in the second equality a fundamental minimax result of game theory permitted us to interchange the max and min operators. This establishes our assertion when substituted in (3.67). □

It is thus clear that the optimality of a design $\xi^\star$ is intimately related to the existence of a certain probability measure $\zeta^\star$ on the σ-algebra of Borel subsets of $\widehat{\Theta}_2(\xi^\star)$. This suggests that the main problem while verifying the optimality of a design is the ease of finding $\zeta^\star$. An advantage in the above formulation is that the problem of verifying the optimality of a design is basically reduced to that of finding a probability measure defined on the σ-algebra of Borel subsets of $\widehat{\Theta}_2(\xi^\star)$. The latter problem is essentially a linear one and therefore should be easier to work with from the computational standpoint. However, except for trivial problems, a general analytic solution for $\zeta^\star$ is usually problematic and one needs to resort to an iterative scheme for determining $\zeta^\star$.

The optimality condition simplifies to a significant measure when $\widehat{\Theta}_2(\xi^\star)$ reduces to a singleton.

Theorem 3.8. *Let Assumptions (A1)–(A3) hold. Given $\xi^\star \in \Xi(X)$, assume that the minimization problem*

$$\min_{\vartheta_2\in\Theta_2} \int_X \|\eta(x) - \eta_2(x,\vartheta_2)\|^2\,\xi^\star(\mathrm{d}x) \tag{3.72}$$

has a unique solution $\vartheta_2^\star \in \Theta_2$. Then:

(i) A necessary and sufficient condition for $\xi^\star$ to be T-optimal is that for each $x \in X$

$$\|\eta(x) - \eta_2(x,\vartheta_2^\star)\|^2 \le \Delta_1(\xi^\star). \tag{3.73}$$

(ii) The equality in (3.73) is attained at all support points of $\xi^\star$, i.e., in the set

$$\operatorname{supp}(\xi^\star) = \bigcap_{A\in\mathcal{B}^+(X)} A, \tag{3.74}$$

where $\mathcal{B}^+(X) = \{A \in \mathcal{B}(X) : \xi^\star(A) > 0\}$.

(iii) the set of all the optimal measures $\xi^\star$ is convex.

Proof. Part *(i)* follows immediately from Theorem 3.7. As for Part *(ii)*, suppose that this were false, i.e., that we could find a set $\widetilde{X} \subset \mathrm{supp}(\xi^\star)$ satisfying $\xi^\star(\widetilde{X}) > 0$ and a scalar $\epsilon > 0$ such that

$$e(x, \vartheta_2^\star) < \Delta_1(\xi^\star) - \epsilon, \quad \forall x \in \widetilde{X}. \tag{3.75}$$

But then we have

$$\begin{aligned} \epsilon \xi^\star(\widetilde{X}) &\leq \int_X [\Delta_1(\xi^\star) - e(x, \vartheta_2^\star)]\, \xi^\star(\mathrm{d}x) \\ &= \Delta_1(\xi^\star) - \int_X e(x, \vartheta_2^\star)\, \xi^\star(\mathrm{d}x) \\ &= \Delta_1(\xi^\star) - \min_{\vartheta_2 \in \Theta_2} \int_X e(x, \vartheta_2)\, \xi^\star(\mathrm{d}x) = 0, \end{aligned} \tag{3.76}$$

which implies $\epsilon \leq 0$, a contradiction.

It remains to show Part *(iii)*, i.e., that the set of optimal measures $\xi^\star$ is convex. But this is immediate, since $\Delta_1(\,\cdot\,)$ is concave. This finishes the proof. □

Note that for an approximate discrete design ξ the definition of $\mathrm{supp}(\xi)$ in (3.74) reduces to that of Definition 2.2. The functions on the left-hand sides of the optimality conditions (here $\psi(x, \xi) = \|\eta(x) - \eta_2(x, \vartheta_2^\star)\|^2$) are sometimes called the *sensitivity functions*. They involve x only and in practice, if $X \subset \mathbb{R}^1$ or $X \subset \mathbb{R}^2$, we can examine their graphs to verify whether or not a given design is optimal. This procedure is illustrated with the following examples.

3.3. Examples

In order to illustrate the use of the Equivalence Theorem in construction of T-optimum designs, we present three examples, where it is possible to obtain explicit expressions defining $\xi^\star$ (otherwise, designs have to be approximated numerically).

Example 3.1. Let the true linear model have the structure

$$\eta(x) = x, \quad x \in X = [-1, 1]. \tag{3.77}$$

We wish to test it against the alternative model

$$\eta_2(x, \vartheta) = \vartheta_2, \quad \vartheta_2 \in [-1, 1] \tag{3.78}$$

which produces a constant response.

Since it is impossible to distinguish between the responses of both the models based on observations at only one support point, suppose that the number of support points is two, i.e., consider prospective designs of the form

$$\xi = \begin{Bmatrix} x_1, & x_2 \\ p_1, & p_2 \end{Bmatrix}. \tag{3.79}$$

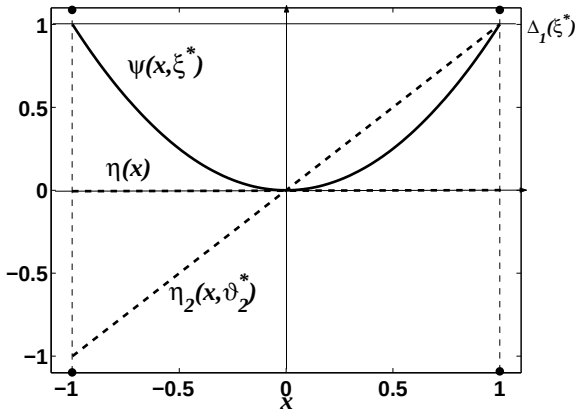


Fig. 3.2. Sensitivity function (solid line) and resultant responses of models (dashed lines) in Example 3.1.

Then

$$\begin{aligned}\mathcal{J}(\xi,\vartheta_2) &= p_1[\eta(x_1)-\eta_2(x_1,\vartheta_2)]^2 + p_2[\eta(x_2)-\eta_2(x_2,\vartheta_2)]^2 \\ &= p_1(x_1-\vartheta_2)^2 + (1-p_1)(x_2-\vartheta_2)^2,\end{aligned} \tag{3.80}$$

where $p_2 = 1 - p_1$ can be eliminated. Therefore

$$\widehat{\vartheta}_2(\xi) = \arg\min_{\vartheta_2} \mathcal{J}(\xi,\vartheta_2) = p_1 x_1 + (1-p_1)x_2. \tag{3.81}$$

Note that $\widehat{\vartheta}_2$ depends on **both the support points and weights**. Furthermore,

$$\Delta_1(\xi) = \mathcal{J}(\xi,\widehat{\vartheta}_2(\xi)) = p_1(1-p_1)(x_1-x_2)^2. \tag{3.82}$$

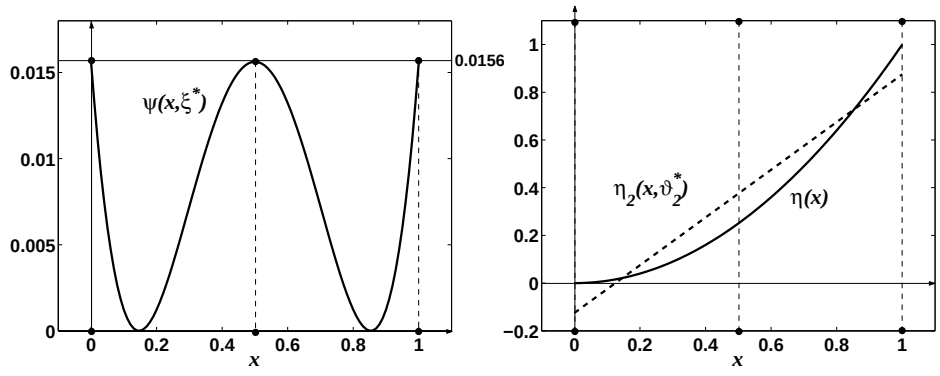
Since the term $p_1(1-p_1)$ is maximal at $p_1 = 1/2$ and $(x_1-x_2)^2$ attains its maximal value at $x_1 = -1$ and $x_2 = 1$, $\Delta_1(\xi)$ is maximized at

$$\xi^\star = \begin{Bmatrix} -1, & 1 \\ 1/2, & 1/2 \end{Bmatrix}. \tag{3.83}$$

Thus we have $\Delta_1(\xi^\star) = 1$, $\widehat{\vartheta}_2(\xi^\star) = 0$ and we get the sensitivity function of the form

$$\psi(x,\xi^\star) = [\eta(x) - \eta_2(x,\widehat{\vartheta}_2(\xi^\star)]^2 = x^2. \tag{3.84}$$

Figure 3.2 shows the sensitivity function and resultant responses of both the models. It can be seen that the situation is consistent with Theorem 3.8, i.e., the global maximum of $\psi(\cdot,\xi^\star)$ over X which equals just $\Delta_1(\xi^\star)$, is attained at support points $x_1^\star = -1$ and $x_2^\star = 1$. Simultaneously, at these support points the square of the difference between the responses of the true and alternative models is maximal.



(a) (b)

Fig. 3.3. Sensitivity function (a) and true and alternative model responses (b) (solid and dashed lines, respectively) in Example 3.2.

Example 3.2. Consider the true model

$$\eta(x) = x^2 \tag{3.85}$$

and the alternative model

$$\eta_2(x, \vartheta_2) = \vartheta_{21} x + \vartheta_{22} \tag{3.86}$$

for the design region $X = [0, 1]$. Note that the number of support points in an optimal discrete design must be at least three, for otherwise, at the worst, it is impossible to distinguish between the responses.

It is an easy exercise to check that the design

$$\xi^\star = \begin{Bmatrix} 0, & 1/2, & 1 \\ 1/4, & 1/2, & 1/4 \end{Bmatrix} \tag{3.87}$$

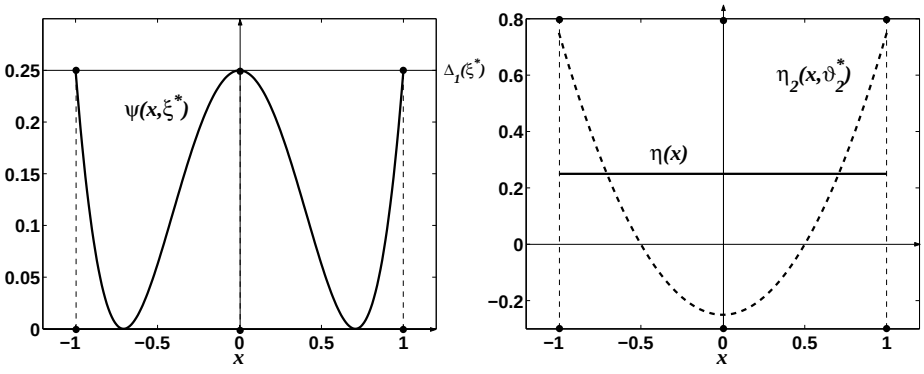
is T-optimal and it corresponds to the unique minimizer $\widehat{\vartheta}_2 = [1, -1/8]^{\mathrm{T}}$. The plot of the corresponding sensitivity function $\psi(x, \xi^\star) = (x^2 - x + 1/8)^2$ is shown in Fig. 3.3 (a). We have $\Delta_1(\xi^\star) = 0.0156$ and this is also the maximal value of $\psi(\cdot, \xi^\star)$. Again, at the support points the square of the difference between the responses of the true and alternative models attains its maximal value, which is shown in Fig. 3.3 (b).

Example 3.3. Consider the true model

$$\eta_1(x) = \gamma, \qquad x \in X = [-1, 1], \tag{3.88}$$

producing a constant response and a quadratic alternative model

$$\eta_2(x, \vartheta_2) = \vartheta_{21} + \vartheta_{22} x + \vartheta_{23} x^2 \tag{3.89}$$



(a) (b)

Fig. 3.4. Sensitivity function (a) and true and alternative model responses for $\gamma = 0.25$ (b) (solid and dashed lines, respectively) in Example 3.3.

with parameter vector $\vartheta_2 = [\vartheta_{21}, \vartheta_{22}, \vartheta_{23}]^{\mathrm{T}}$.

In the absence of any constraints on the parameters of the second model, if the first model is true, so is the second and designing an experiment for detecting departures from Model 2 is meaningless. Instead, we employ the constraints

$$\Theta_2 = \left\{(\vartheta_{21}, \vartheta_{22}, \vartheta_{23}) : \vartheta_{22}^2 + \vartheta_{23}^2 \geq 1, \quad \vartheta_{22}, \vartheta_{23} \geq 0\right\}. \tag{3.90}$$

Therefore

$$\begin{aligned} \Delta_1(\xi) &= \min_{\vartheta_2 \in \Theta_2} \mathcal{J}(\xi, \vartheta_2) \\ &= \min_{\vartheta_2 \in \Theta_2} \int_X \left(\gamma - (\vartheta_{21} + \vartheta_{22} x + \vartheta_{23} x^2)\right)^2 \xi(\mathrm{d}x). \end{aligned} \tag{3.91}$$

The minimum of the above non-centrality parameter for a given design ξ evidently occurs on the subset of Θ_2 for which $\vartheta_{22}^2 + \vartheta_{23}^2 = 1$.

We will find an optimum design within a class of symmetric three-point designs and then use Theorem 3.8 to check that this design is T-optimum amongst all designs, not just within the restricted class.

Suppose that the design has the form

$$\xi(p) = \begin{Bmatrix} -1, & 0, & 1 \\ p/2, & 1-p, & p/2 \end{Bmatrix}. \tag{3.92}$$

We are looking for an optimum value of the parameter p.

We have

$$\begin{aligned}\widehat{\vartheta}_2 &= \arg\min_{\vartheta_2\in\Theta_2} \mathcal{J}(\xi(p),\vartheta_2) = \arg \min_{\substack{\vartheta_{21}\\ \vartheta_{23}=\sqrt{1-\vartheta_{22}^2}}} \Big(\frac{p}{2}(\gamma-\vartheta_{21}+\vartheta_{22}-\vartheta_{23})^2 \\ &\quad + (1-p)(\gamma-\vartheta_{21})^2 + \frac{p}{2}(\gamma-\vartheta_{21}-\vartheta_{22}-\vartheta_{23})^2\Big) \\ &= \arg\min_{\substack{\vartheta_{21}\\ 0\le\vartheta_{22}\le 1}} \left(\frac{p}{2}\left(\gamma-\vartheta_{21}+\vartheta_{22}-\sqrt{1-\vartheta_{22}^2}\right)^2\right. \\ &\quad \left. + (1-p)(\gamma-\vartheta_{21})^2 + \frac{p}{2}\left(\gamma-\vartheta_{21}-\vartheta_{22}-\sqrt{1-\vartheta_{22}^2}\right)^2\right) \\ &= \arg\min_{\substack{\vartheta_{21}\\ 0\le\vartheta_{22}\le 1}} \left(p\left(\gamma-\vartheta_{21}-\sqrt{1-\vartheta_{22}^2}\right)^2 + p\vartheta_{22}^2 + (1-p)(\gamma-\vartheta_{21})^2\right).\end{aligned} \tag{3.93}$$

Finally, we obtain

$$\widehat{\vartheta}_2 = [\widehat{\vartheta}_{21},\widehat{\vartheta}_{22},\widehat{\vartheta}_{23}]^{\mathrm{T}} = [\gamma-p,0,1]^{\mathrm{T}} \tag{3.94}$$

and

$$\Delta_1(\xi(p)) = \mathcal{J}(\xi(p),\widehat{\vartheta}_2(\xi)) = p(1-p). \tag{3.95}$$

It is obvious that $\Delta_1(\xi(p))$ is maximized for $p=1/2$ and the optimum design within the restricted class has the form

$$\xi^\star = \begin{Bmatrix} -1, & 0, & 1 \\ 1/4, & 1/2, & 1/4 \end{Bmatrix}. \tag{3.96}$$

Thus we have $\Delta_1(\xi^\star)=1/4$ and we get the sensitivity function in the form

$$\psi(x,\xi^\star) = (1/2-x^2)^4. \tag{3.97}$$

Figure 3.4 depicts it and shows again that the found design is definitely the T-optimal one. Also the responses of both the models for $\gamma=1/4$ are shown.

3.4. Summary

In this chapter, the T-optimality criterion has been defined and characterized. Conceptually, for linear models it is proportional to the non-centrality parameter of the χ^2 distribution of the residual sum of squares for the second model. A T-optimum design leads to the most powerful F-test for lack of fit of the second model when the first is true. When the models are nonlinear in the parameters, the exact F-test is replaced by asymptotic results.

This seemingly simple criterion becomes non-differentiable as soon as the minimum squared difference between the true and predicted responses integrated with respect to the current design is attained at more than one parameter value ϑ_2. However, the situation is not that hopeless, as we have shown a number

of characterizations of $\Delta_1(\,\cdot\,)$ and the corresponding specialized version of the equivalence theorem. The derivation of those results was made possible at the cost of using quite advanced mathematical notions, such as weak convergence of probability measures, but this is an unavoidable price to be paid if we attempt to attack the problem properly from the mathematical point of view.

In addition to revealing striking minimax properties of T-optimal designs, the Equivalence Theorem which is our main result in this chapter provides us with a test for the T-optimality of intuitively sensible designs. However, to exploit its full potential, we require more that this, namely, we need efficient numerical algorithms which would enable us to construct T-optimum design measures. Note that the problem is not that standard if we cannot initially identify a finite set which includes the support points of an optimal measure. In fact, the Equivalence Theorem says nothing about the number of the support points of an optimal design and no bound on this number is known in the literature. What is more, there is no guarantee that an optimal design is unique. In particular, multiple global solutions $\xi^\star$ may yield the same maximizing value of $\Delta_1(\xi)$. Furthermore, there may be multiple local optima of functions involved in construction of optimum designs, which highly interferes with the optimization process. Consequently, the crucial question is: How to efficiently generate approximations to T-optimum designs? Unfortunately, it appears that no known specialized algorithm exists for constructing T-optimum designs, apart from an adaptation of the Wynn-Fedorov scheme (Atkinson and Fedorov, 1975a; Fedorov and Hackl, 1997) which suffers from serious problems with convergence. This computational problem is far from being satisfactory solved and the remainder of the dissertation is intended as an attempt to fill this gap.

Chapter 4

ALGORITHMIC METHODS FOR T-OPTIMUM EXPERIMENTAL DESIGN

In addition to revealing intriguing minimax properties of T-optimal designs, the equivalence theorems of Chapter 3 provide us with tests for the T-optimality of intuitively sensible designs. To be specific, consider a design ξ for which the answering set reduces to a singleton, i.e.,

$$\widehat{\Theta}_2(\xi) = \operatorname{Arg} \min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x) = \{\widehat{\vartheta}_2\} \tag{4.1}$$

for some $\widehat{\vartheta}_2 \in \Theta_2$. The T-optimality of ξ can be examined as follows:

1 If the sensitivity function $\psi(x, \xi) = \|\eta(x) - \eta_2(x, \widehat{\vartheta}_2)\|^2$ is less than or equal to $\Delta_1(\xi)$ for all $x \in X$, then ξ is T-optimal.

2 If the sensitivity function $\psi(x, \xi)$ exceeds $\Delta_1(\xi)$, then ξ is not T-optimal.

However, in order to exploit fully the salient features of the optimality conditions, we need efficient computational methods which would enable us to construct T-optimum design measures. In this chapter we consider three such methods and include a thorough discussion of their effectiveness. We remark that the construction of a solution to the T-optimum design problem is not straightforward, owing to the inherent nondifferentiability of the optimality criterion when the answering set is not a singleton.

Specifically, the adaptation of the classical Wynn-Fedorov algorithm is considered in Section 4.1. In Section 4.2 we solve the original problem by solving directly the associated semi-infinite optimization problem. Then, in Section 4.3 we propose a relaxation algorithm which constitutes one of the most important original contributions made in the dissertation to the development of numerical algorithms for the T-optimum design problem. In the remainder of this chapter, an auxiliary problem of selecting global optimizers is discussed. Thus, Section 4.4 outlines a possible choice of the global optimization method being an essential component of the proposed computational schemes. Chapter 4 concludes with the consideration of several illustrative examples.

4.1. Wynn-Fedorov algorithm for constructing T-optimal designs

An iterative algorithm of the Wynn-Fedorov type outlined in Section 2.4 was recommended by (Atkinson and Fedorov, 1975a) to construct iteratively T-optimum designs for static single-output systems (Atkinson and Fedorov, 1975a).

Based on the Equivalence Theorem, (cf. Theorem 3.8), this algorithm can be directly adopted to multi-output dynamic systems. The corresponding numerical scheme can be represented by the following steps (n^k stands for size of the design in k-th step):

Algorithm 4.1 (Generalized Wynn-Fedorov algorithm for T-optimum designs)**.**

Step 1: Guess an initial design ξ^0 of the form $\xi^0 = \begin{Bmatrix} x_1^0, & \dots, & x_{n^0}^0 \\ p_1^0, & \dots, & p_{n^0}^0 \end{Bmatrix}$ for some arbitrary n^0. Choose some positive tolerance $\epsilon \ll 1$. Set $k = 0$.

Step 2: Determine

$$\begin{aligned} \widehat{\vartheta}_2^k &= \arg \min_{\vartheta_2 \in \Theta_2} \sum_{i=1}^{n^k} p_i^k \|\eta(x_i^k) - \eta_2(x_i^k, \vartheta_2)\|^2, \\ \widehat{x}^k &= \arg \max_{x \in X} \|\eta(x) - \eta_2(x, \widehat{\vartheta}_2^k)\|^2. \end{aligned} \tag{4.2}$$

Step 3: If $\psi(\widehat{x}^k, \xi^k) \le \Delta_1(\xi^k) + \epsilon$, where

$$\begin{aligned} \psi(x, \xi^k) &= \|\eta(x) - \eta_2(x, \widehat{\vartheta}_2^k)\|^2, \\ \Delta_1(\xi^k) &= \sum_{i=1}^{n^k} p_i^k \|\eta(x_i^k) - \eta_2(x_i^k, \widehat{\vartheta}_2^k)\|^2, \end{aligned} \tag{4.3}$$

then $\xi^\star = \xi^k$ and STOP.

Step 4: Choose the appropriate value of α^k, $0 \le \alpha^k \le 1$, and compute the convex combination of designs:

$$\xi^{k+1} = (1 - \alpha^k)\xi^k + \alpha^k \delta(\widehat{x}^k), \tag{4.4}$$

where $\delta(\widehat{x}^k)$ is the design concentrated only at one support point $\widehat{x}^k$.
Set $k \leftarrow k + 1$ and go to Step 2.

A slight complication compared with the case of a single-response system consists in the necessity of calculation of the Euclidean distance between the responses instead of the absolute value. In contrast to the D-optimum counterpart, even in the single response case, selection of an optimum value of the steplength α^k for $\Delta_1(\xi^{k+1}, \widehat{\vartheta}_2^k)$, i.e., finding

$$\alpha^k = \arg \max_{\alpha \in [0,1]} \Delta_1((1 - \alpha)\xi^k + \alpha \delta(\widehat{x}^k)),$$

necessitates a numerical search. To this end (since $\alpha \in [0,1]$) a simple search procedure, e.g., the golden section method (Press *et al.*, 1996) can be used. Alternatively, like in the D-optimum case, another common choice is

$$\alpha^k = \frac{1}{k+1} \tag{4.5}$$

or, in general, α^k can be chosen using any sequence satisfying the conditions (Fedorov and Hackl, 1997)

$$\lim_{k\to\infty} \alpha^k = 0, \quad \sum_{k=0}^{\infty} \alpha^k = \infty, \quad \sum_{k=0}^{\infty} (\alpha^k)^2 < \infty. \tag{4.6}$$

Generally, the convergence speed of the presented scheme is rather low, since it actually belongs to the group of first-order algorithms. In practice, the optimum support points usually are found relatively quickly (when using efficient global optimizers during Step 2 of the algorithm), but a precise determination of the corresponding weights takes much more time. In the literature there have been some attempts to modify the basic scheme so as to enhance the convergence speed (Fedorov and Hackl, 1997). The resulting heuristics are intended mainly for the D-optimum design criterion, but with minor changes they can be adapted to the T-optimum criterion.

For example, one of the characteristic features of the algorithm is that the weights of the non-optimal support points are gradually decreased. This eventually results in the existence of support points with negligible weights. Moreover, in each iteration a new support point is additionally included into the design. Usually, after several iterations, the location of the new points becomes similar or very close to the existing ones. This is caused by the numerical inaccuracies of the optimization process. In order to obtain minimal-support solutions it is worthwhile to equip the implementation with procedures aimed at removing support points with negligible weights from the current design and replacing clustered points by a single support point. The weights of the points to be replaced are added in each cluster and the clustered points are substituted by only one point with the weight equal to the resulting sum. Then removing points with negligible weights can be performed. The thresholds defining a maximal radius of the clusters (and, consequently, the number of the replaced points) and a minimum acceptable weight are parameters which should be set *a priori*. The appropriate choice can speed up the convergence, but it may happen that by setting excessively high thresholds we will obtain the effect of repeatedly removing and adding the same points into the design (and, consequently, the lack of the convergence of the entire scheme). In the case of D-optimality a useful test exists (Pronzato, 2003) which allows for a safe removal of points which have no chance to be located in the optimum design. Unfortunately, there is no such counterpart here and the appropriate thresholds should be chosen empirically.

For dynamic systems described by ordinary differential equations we can directly use the presented algorithm while interpreting x as time. Application to

systems with spatio-temporal dynamics described by partial differential equations is additionally complicated by the necessity of additional integration with respect to the time variable (Kuczewski, 2005). This topic will be discussed in the Chapter 6.

The presented computational scheme still constitutes a basis for practitioners. However, it suffers from some impediments outlined above. It seems that the most confounded nuisance which limits applications consists in a possible lack of convergence when non-unique minimizers $\widehat{\vartheta}_2^k$ exist for a given ξ^k. This phenomenon should not be surprising, since using the formulae for both the maximizer $\widehat{x}^k$ in Step 2, and the sensitivity function $\psi(x, \xi^k)$ in Step 3, we take implicitly for granted that the answering set $\widehat{\Theta}_2(\xi^k) = \operatorname{Arg} \min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi^k(\mathrm{d}x)$ is a singleton. To see this, it is sufficient to compare the form of the stopping criterion in Step 3 with the optimality conditions of Theorem 3.8. A detailed proof of the convergence, from which it would be clearer, is not provided here (by the way, to the best of the author's knowledge, it does not seem to be published, at least in the English literature), but it proceeds on the same lines as that for the algorithm dedicated to the DT-optimality criterion, cf. Chapter 5. The form of the optimality conditions for the general case is rather awkward, cf. Theorem 3.7, and consequently, they can hardly be used to alter the foregoing exchange algorithm so as to make it reliable. This was one of the main motivations to write this dissertation.

Example 4.1. In this contrived example, consider the true model having the structure

$$\eta(x) = \sin(x) + \cos(2x), \quad x \in X = [-1, 1]. \tag{4.7}$$

We wish to test it against the alternative quadratic model which is linear with respect to the unknown parameters:

$$\eta_2(x, \vartheta_2) = \vartheta_{20} + \vartheta_{21} x + \vartheta_{22} x^2. \tag{4.8}$$

Starting with a randomly chosen three-point uniformly distributed (i.e., with all weights equal to 1/3) initial design and the numerical accuracy set to $\epsilon = 0.0002$, Algorithm 4.1 found in 100 iterations the four-point T-optimal design of the following form (the values were rounded to the third fractional digit and some additional modifications of the basic procedure, like a removal of clusters and points with negligible weights, were also applied):

$$\xi^\star = \begin{Bmatrix} -1.000, & -0.673, & 0.071, & 0.827 \\ 0.257, & 0.472, & 0.247, & 0.023 \end{Bmatrix}. \tag{4.9}$$

The relevant computer program was implemented using Maple 10 as a set of Maple procedures. Figure 4.1 displays the resultant sensitivity function $\psi(x, \xi^\star)$ as a function of the x variable ($\vartheta_2^\star = [\vartheta_{20}^\star, \vartheta_{21}^\star, \vartheta_{22}^\star]^{\mathrm{T}} = [0.913, 0.909, -1.349]^{\mathrm{T}}$). Again, it can be observed that the situation is consistent with the developed equivalence theorem which confirms the optimality of the found approximation (up to the assumed numerical accuracy ϵ).

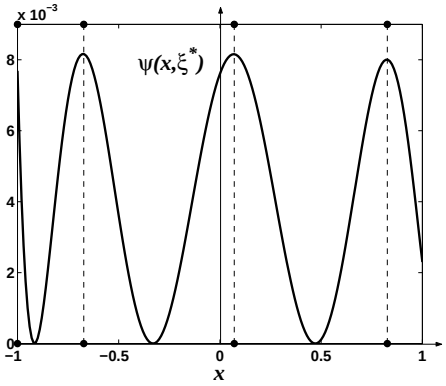


Fig. 4.1. Sensitivity function in Example 4.1.

Example 4.2. Consider two types of chemical reactions which turn substance A into C and proceed with creation of indirect reagent B. The first model is called the consecutive irreversible reaction A $\rightarrow$ B $\rightarrow$ C and is described by the system of ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A]^{\lambda_1}, \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A]^{\lambda_1} - k_2[B]^{\lambda_2}, \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B]^{\lambda_2}, \end{aligned} \tag{4.10}$$

with initial conditions

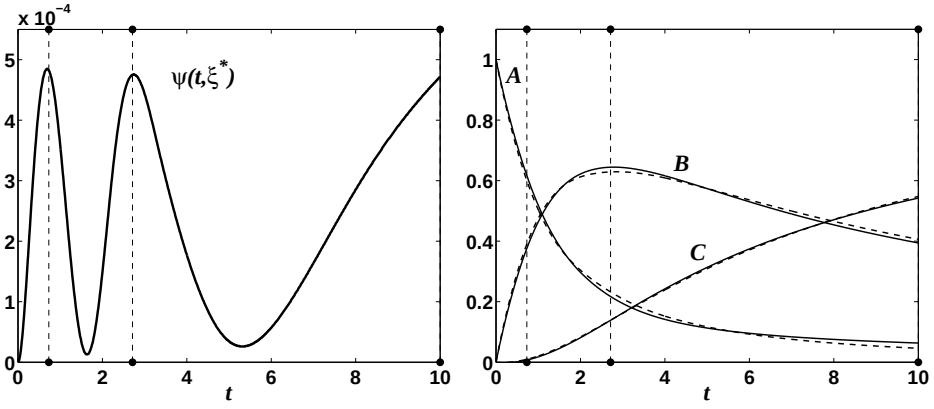
$$[A]_{t=0} = a_0, \quad [B]_{t=0} = b_0, \quad [C]_{t=0} = c_0,$$

where $[A]$, $[B]$, $[C]$ denote the concentrations of the reagents A, B and C, respectively, a_0, b_0 and c_0 being the corresponding initial concentrations. Such a scheme is typical, e.g., for the phenomenon of radioactive decay. Another type of reaction is described by the system of ODE's

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A]^{\lambda_1} + k_3[B]^{\lambda_3}, \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A]^{\lambda_1} - k_2[B]^{\lambda_2} - k_3[B]^{\lambda_3}, \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B]^{\lambda_2}, \end{aligned} \tag{4.11}$$

with initial conditions

$$[A]_{t=0} = a_0, \quad [B]_{t=0} = b_0, \quad [C]_{t=0} = c_0.$$



(a) (b)

Fig. 4.2 . Sensitivity function (a) and responses of true and alternative models (b) (continuous and dashed lines, respectively) obtained in Example 4.2. The dashed vertical lines correspond to the location of two optimal support points (the third point is just the end of the observation horizon).

These are kinetic equations of a reversible first-order reaction followed by an irreversible reaction $\mathrm{A} \rightleftharpoons \mathrm{B} \rightarrow \mathrm{C}$, where the transitional substance B achieves an initial balance with reagent A.

We assume that the true model is described by (4.11) for the setting

$$\vartheta_1 = (k_1, k_2, k_3, \lambda_1, \lambda_2, \lambda_3) = (0.7, 0.2, 0.1, 2.0, 2.0, 1.0).$$

The reaction described by (4.10) constitutes the alternative model. The purpose of the design consists in determining a T-optimal schedule (here, time instances) of measurements of reagent concentrations which will allow us to maximize the certainty of the discrimination between the above two competing models. The design range and initial concentrations were respectively set to $T = [0, 10]$ and $(a_0, b_0, c_0) = (1, 0, 0)$ for both the models. The set of feasible values of the alternative model parameters was set to $0.55 \leq k_1 \leq 0.85$, $0.05 \leq k_2 \leq 0.35$, $1.5 \leq \lambda_1 \leq 2.5$, $1.5 \leq \lambda_2 \leq 2.5$. The program implementing Algorithm 4.1 was written in Lahey-Fujitsu Fortran 95 compiler v.5.6 using the IMSL library. In much the same way as in Example 4.1, additional modifications of the basic procedure were applied. To solve the implied global optimization problems, the Adaptive Random Search (ARS) method was utilized, cf. Section 4.4. The resultant design has the form

$$\xi^\star = \begin{Bmatrix} 0.73, & 2.71, & 10.00 \\ 0.194, & 0.365, & 0.441 \end{Bmatrix}. \tag{4.12}$$

This means that, e.g., about one fifth of the whole experimental effort should be concentrated at time instant $t = 0.73$. The least profitable values of the alternative model parameters with respect to the T-optimum design $\xi^\star$ are $\vartheta_2^\star =$

$(k_1, k_2, \lambda_1, \lambda_2) = (0.847, 0.198, 1.575, 1.981)$. For $\epsilon = 0.00001$ convergence was achieved in 89 iterations. The plot of the obtained sensitivity function $\psi(t, \xi^\star) = \|\eta(t) - \eta_2(t, \vartheta_2^\star)\|^2$ is shown in Fig. 4.2. The same figure illustrates also the responses of both the models.

4.2. Semi-infinite programming

The problem of finding T-optimum designs can be transformed to the so-called Semi-Infinite Programming one (SIP) and then efficiently solved using one of the dedicated methods. That approach has one decided advantage over the Wynn-Fedorow scheme, namely, the theory in SIP, as well as the numerical SIP methods have expanded very fast during the last years (Reemtsen and Rückmann, 1998). In turn, a serious drawback is that the knowledge of an upper bound on the number of support points in the optimal design is required, which may be problematic in applications. Nevertheless, the approach is attractive, especially when the Wynn-Fedorov algorithm fails.

A typical semi-infinite programming problem has the following form (Žaković and Rustem, 2003; Hettich and Kortanek, 1993):

$$\begin{aligned} &\min_{u \in U} f(u) \\ &G(u, v) \leq 0, \quad \forall v \in V, \end{aligned} \tag{4.13}$$

where $u \in U \subset \mathbb{R}^r$, $v \in V \subset \mathbb{R}^\rho$, $f \in C^2(U)$, $G \in C^2(U \times V; \mathbb{R}^{r+\rho})$, U and V being some nonempty compact sets. The term semi-infinite programming derives from the fact that although we have a finite number of decision variables in the target function f, the condition $G(u, v) \leq 0,\ \forall\, v \in V$ represents an infinite set of constrains on u.

Interest in the SIP problem dates back to the late 1940 and since then it has been extensively studied in the literature (the reader is referred, e.g., to the survey (Hettich and Kortanek, 1993) and the references therein). Applications of SIP include, among other things, minimax and saddle point problems. Also, their applications to numerous problems in science and engineering are extensive (e.g., some problems in mechanics, environmental sciences or engineering design, cf. (Polak, 1987)).

Popular numerical processing tools such as MATLAB often include ready-to-use procedures for solving SIP problems. As an example, consider the MATLAB *fseminf* procedure included in MATLAB Optimization Toolbox, which is capable of solving small-scale SIP problems (MathWorks, 2005).

4.2.1. T-optimum design as an SIP problem

It is well known (Žaković and Rustem, 2003), that a typical minimax problem

$$\min_{u \in U} \max_{v \in V} f(u, v) \tag{4.14}$$

can alternatively be expressed as the problem

$$\left\{\begin{array}{ll} & \min\limits_{u\in U,\alpha}\ \alpha \\ \text{subject to} & \max\limits_{v\in V} f(u,v)\le\alpha, \end{array}\right. \tag{4.15}$$

which finally leads to some SIP problem

$$\left\{\begin{array}{ll} & \min\limits_{u\in U,\alpha}\ \alpha \\ \text{subject to} & f(u,v)-\alpha\le 0,\quad \forall\, v\in V. \end{array}\right. \tag{4.16}$$

Such a transformation allows us to handle a large class of minimax problems arising in numerous disciplines (e.g., engineering, finance, decision making under uncertainty) as SIP problems.

Also the original problem (3.14) of T-optimum design

$$\Delta_1(\xi^\star)=\max_{\xi\in\Xi(X)}\ \min_{\vartheta_2\in\Theta_2}\ \mathcal{J}(\xi,\vartheta_2), \tag{4.17}$$

where $\mathcal{J}(\xi,\vartheta_2)=\int_X \|\eta(x)-\eta_2(x,\vartheta_2)\|^2\,\xi(\mathrm{d}x)$, can be similarly transformed into an 'SIP-like' formulation

$$\left\{\begin{array}{ll} & \max\limits_{\xi\in\Xi(X),\alpha}\ \alpha \\ \text{subject to} & \mathcal{J}(\xi,\vartheta_2)\ge\alpha,\quad \forall\vartheta_2\in\Theta_2, \end{array}\right. \tag{4.18}$$

or equivalently,

$$\left\{\begin{array}{ll} & \min\limits_{\xi\in\Xi(X),\beta}\ \beta \\ \text{subject to} & -\beta-\mathcal{J}(\xi,\vartheta_2)\le 0,\quad \forall\vartheta_2\in\Theta_2, \end{array}\right. \tag{4.19}$$

which fits snugly into the general problem (4.13) if the number of the support points in ξ is fixed *a priori*.

Indeed, assuming that a T-optimal design has the discrete form

$$\xi=\begin{Bmatrix} x_1, & \ldots, & x_n \\ p_1, & \ldots, & p_n \end{Bmatrix}$$

for a fixed n which should be chosen as a sufficiently large number (in practice, we may only guess this value, as no theoretical results exist regarding the number of support points in T-optimal designs), we can define

$$\begin{aligned} U&=\left\{(x_1,\ldots,x_n,p_1,\ldots,p_n,\beta)\in X^n\times[0,1]^n\times[-c,0]:\sum_{i=1}^n p_i=1\right\}\subset\mathbb{R}^{ns+n+1}\\ V&=\Theta_2\subset\mathbb{R}^m, \end{aligned} \tag{4.20}$$

where c is a sufficiently large positive number, and write

$$\begin{aligned} f(u)&=\beta,\\ G(u,v)&=-\beta-\sum_{i=1}^n p_i\|\eta(x_i)-\eta_2(x_i,\vartheta_2)\|^2. \end{aligned} \tag{4.21}$$

This means that numerical methods for SIP can be directly applied to find approximations to the sought T-optimal designs.

4.2.2. Numerical methods for SIP

Generally, numerical solution of SIP problems usually consists in generating a sequence of finitely constrained auxiliary optimization problems suitable for being solved by standard algorithms for finite constrained optimization. Depending on the way in which those auxiliary problems are generated, the following three classes of applicable algorithms can be distinguished (Hettich *et al.*, 2001):

- exchange methods (including cutting plane methods which are applicable only for convex problems),
- discretization methods,
- methods based on local reduction.

In order to present the scheme of the exchange and discretization methods, we rewrite the initial SIP problem (4.13) in the form

$$\mathcal{S}(V)\begin{cases} \min\limits_{u\in\widetilde{U}} f(u) \\ \widetilde{U} = \Big\{u \in U : G(u,v) \le 0,\ \forall y \in Y\Big\}. \end{cases} \tag{4.22}$$

Problem $\mathcal{S}(V)$ is then approximated by finitely constrained auxiliary problems $\mathcal{S}(V^k)$ in the form

$$\mathcal{S}(V^k)\begin{cases} \min\limits_{u\in U^k} f(u) \\ U^k = \Big\{u \in U^0 : G(u,v) \le 0,\ \forall v \in V^k\Big\}, \end{cases} \tag{4.23}$$

where U^0 is a convex compact set, $U^0 \subset U$ and $V^k \subset V$ is a set containing a finite number of elements.

Before we get into details of the above-mentioned methods, it is worth of noticing that there also exists a different approach to solving SIP problems. Namely, by transforming the constraints $G(u,v) \le 0,\ \forall v \in V$ into $\max\limits_{v\in V} G(u,v) \le 0$, the problem may be treated in the framework of non-differentiable optimization. The interested reader is referred to the survey paper (Polak, 1987).

In the sequel, we assume that both f and G are smooth and twice continuously differentiable functions.

4.2.2.1. Exchange methods

The name of the whole family of such methods originates from the fact that during the k-th iteration of the procedure, an exchange of constraints takes place, i.e., the set V^{k+1} is obtained from V^k by addition of at least one new constraint and (usually) deletion of some of the constraints, i.e., elements of the V^k set.

The general scheme of an exchange procedure can be described by the following steps (Hettich and Kortanek, 1993):
During the k-th iteration

❶ For given $V^k \subset V$ find a solution $\widehat{u}^k$ of the problem $\mathcal{S}(V^k)$.

❷ Find some (or all) maxima $\widehat{v}_1^k, \ldots, \widehat{v}_{q_k}^k$ of the subproblem $\mathcal{Q}(\widehat{u}^k)$ defined as follows:

$$\mathcal{Q}(\widehat{u}^k): \quad \max_{v \in V} G(\widehat{u}^k, v). \tag{4.24}$$

❸ If

$$G(\widehat{u}^k, \widehat{v}_j^k) \leq 0, \quad j = 1, \ldots, q_k, \tag{4.25}$$

then STOP ($\widehat{u}^k$ is the optimum solution), otherwise construct V^{k+1} such that

$$V^{k+1} \subset V^k \cup \{\widehat{v}_1^k, \ldots, \widehat{v}_{q_k}^k\}, \tag{4.26}$$

set $k \leftarrow k+1$ and proceed to the next iteration.

The condition necessary for convergence of the scheme is

$$\max_{j=1,\ldots,q_k} G(\widehat{u}^k, \widehat{v}_j^k) = \max_{v \in V} G(\widehat{u}^k, v) \tag{4.27}$$

i.e., we need to find a global optimum of the subproblem $\mathcal{Q}(\widehat{u}^k)$ which for higher dimensions of the V set may be very costly.

A lot of realizations of the presented scheme exist in the literature (see, e.g., references in (Hettich *et al.*, 2001)), including suggestions to use the cutting plane algorithm for linear problems (Wu *et al.*, 1998) and convex problems (Kortanek and No, 1993). The particular realizations differ mainly in a manner of constructing the set V^{k+1}, e.g., in (Hettich and Kortanek, 1993) the convergence of the above scheme is proved in the case when no constraint is deleted from V^k.

It seems that the simplest implementation of the exchange method consists in keeping all old constraints and adding only one new constraint in each iteration (i.e., $V^{k+1} \subset V^k \cup \{\widehat{v}^k\}$) satisfying $\widehat{v}^k = \max_{v \in V} G(\widehat{u}^k, v)$. In the next section we adapt this procedure to solve the T-optimum SIP problem (4.19).

4.2.2.2. Discretization methods

Methods belonging to this family generate a finite number of constraints by scattering the V set using a grid which is usually regular. Consequently, a sequence of problems $\mathcal{S}(V^k)$ is generated, where V^k is an h^k-grid on V such that

$$\sup_{v \in V} \operatorname{dist}(v, V^k) \leq h^k, \tag{4.28}$$

where $h^k > 0$ defines the maximal stepsize over all the dimensions.

Since the grid can generate a huge number of constraints, due to efficiency reasons the key idea is to choose in the k-th iteration only a subset $\bar{V}^k$ of the

whole V^k. The set $\bar{V}^k$ is consequently enlarged as long as the solution $\widehat{u}^k$ of the problem $\mathcal{S}(\bar{V}^k)$ is feasible also for V^k. Then, the grid is refined: $h^{k+1} = \gamma^k h^k$ with $\gamma^k \in (0,1)$, and the next iteration proceeds. Such a scheme is repeated until a prescribed number of iterations is achieved.

Generally, the algorithm can be described by the following steps:
During the k-th iteration, given h^{k-1}, the last set $\bar{V}^{k-1} \subset V^{k-1}$ and the solution $\widehat{u}^{k-1}$ to the problem $\mathcal{S}(\bar{V}^{k-1})$

❶ Choose $h^k = \gamma^k h^{k-1}$ and generate V^k,

❷ Select $\bar{V}^k \subset V^k$,

❸ Compute a solution $\bar{u}^k$ to $\mathcal{S}(\bar{V}^k)$. If $\bar{u}^k$ is also feasible for $\mathcal{S}(V^k)$ within a given accuracy, then set $\widehat{u}^k = \bar{u}^k$ and go to Step 4. Otherwise, repeat Step 2 for a new choice of $\bar{V}^k$, enlarging the old one,

❹ If $k > k_{max}$, where k_{max} is an arbitrarily chosen number of grid refinement steps, then STOP ($\widehat{u}^k$ is the optimum solution). Otherwise, set $k \leftarrow k+1$ and go to Step 1 (repeat the procedure again).

An important aspect of that type of algorithms as regards efficiency is the use of as much information as possible from the previous grids while solving $\mathcal{S}(\bar{V}^k)$. Implementations differ mainly in the choice of $\bar{V}^k$, but a common suggestion is to select

$$\bar{V}^k \supset V^k_\varsigma = \{v \in V^k : G(\widehat{u}^k, v) \geq -\varsigma\} \tag{4.29}$$

with some threshold $\varsigma > 0$ selected a priori. The choice of ς is ambiguous. Its too large value leads to many constraints in $\mathcal{S}(\bar{V}^k)$, which decreases efficiency. However, a too small ς may lead to multiple repetition of the second and third steps of the algorithm during each iteration. There exist adaptive strategies of gradually decreasing ς (Painter and Tits, 1989). Also the coefficient γ^k responsible for the grid refinement can be chosen in many ways (e.g., based on the information about last setting $\bar{V}^{k-1}$ and solution $\widehat{u}^{k-1}$ (Polak and He, 1992)).

4.2.2.3. Local reduction methods

The methods of this type are founded on the assumption that, under appropriate conditions, the original continuum of constraints can be replaced by finitely many constraints which are locally sufficient to describe the feasible solution region U. Thus, against the original one, the reduced problem is being solved in each iteration. The basis for that reduction can be presented as follows (Hettich *et al.*, 2001):

Let the f and G be smooth and twice continuously differentiable functions ($f \in C^2(\mathbb{R}^r)$, $G \in C^2(\mathbb{R}^r \times V)$), $\bar{u} \in \mathbb{R}^r$ be a given point and $\bar{v}_1, \ldots, \bar{v}_{q(\bar{u})}$, $q(\bar{u}) < \infty$, be all local solutions to the problem $\mathcal{Q}(\bar{u}) : \max_{v \in V} G(\bar{u}, v)$. It is obvious that $\bar{u}$ is feasible *iff* $G(\bar{u}, \bar{v}_j) \leq 0$, $j = 1, \ldots, q(\bar{x})$. Furthermore, assume that there exist a neighborhood $\bar{\mathcal{N}}$ of $\bar{u}$ and twice continuously differentiable functions:

$$v_j : \bar{\mathcal{N}} \to V, \quad v_j(\bar{u}) = \bar{v}_j, \quad j = 1, \ldots, q(\bar{u}) \tag{4.30}$$

such that for every $u \in \bar{\mathcal{N}}$ we have that $v_j(u)$ are all local solutions of $\mathcal{Q}(u)$. Then, with

$$G_j(u) = G(u, v_j(u)), \quad j = 1, \ldots, q(\bar{u}), \tag{4.31}$$

we have

$$G_j \in C^2(\bar{\mathcal{N}}) \tag{4.32}$$

and

$$U \cap \bar{\mathcal{N}} = \{u \in \bar{\mathcal{N}} : G_j(u) \leq 0, \quad j = 1, \ldots, q(\bar{u})\}. \tag{4.33}$$

This means that in $\bar{\mathcal{N}}$ the original problem $\mathcal{S}(V)$ can be replaced by the reduced finite one in the form

$$\mathcal{S}_{\bar{u}}(V) \begin{cases} \min\limits_{u \in U^0} f(u) \\ U^0 = \Big\{u \in U : G_j(u) \leq 0, \quad j = 1, \ldots, q(\bar{u})\Big\}. \end{cases} \tag{4.34}$$

The general scheme of this type of algorithm can be expressed by the following steps:
During the k-th iteration, given $\widehat{u}^{k-1}$ (not necessarily feasible),

❶ Find **all** local maxima $\widehat{v}_1^k, \ldots, \widehat{v}_{q_k}^k$ of the subproblem $\mathcal{Q}(\widehat{u}^k)$ defined as follows:

$$\mathcal{Q}(\widehat{u}^k): \quad \max_{v \in V} G(\widehat{u}^k, v) \tag{4.35}$$

❷ Apply some algorithm of finite programming to get the solution u^k of the *reduced* problem $\mathcal{S}^k$,

$$\mathcal{S}^k \begin{cases} \min\limits_{u \in U^k} f(u) \\ U^k = \Big\{u \in U : G_j(u) \leq 0, \quad j = 1, \ldots, q_k\Big\}, \end{cases} \tag{4.36}$$

where

$$G_j(u) = G(u, \widehat{v}_j(u)) \tag{4.37}$$

and the functions $\widehat{v}_j(u)$ are defined in a neighbourhood of $\widehat{u}^{k-1}$.

❸ Set $\widehat{u}^k \leftarrow u^k$, $k \leftarrow k+1$ and proceed to the next iteration.

Computationally, Step 1 is the most costly since it requires a global search for all maxima of $\mathcal{Q}(\widehat{u}^k)$. Hence in practical realizations the execution of this step should be avoided as much as possible. A basic assumption of the reduction method is that there exist a finite number of such maxima. If it fails, the method is not applicable. Step 2 necessitates the use of some method of finite non-linear programming. Effective Sequential Quadratic Programming (SQP) methods have been widely used in this context, especially the ones using the augmented Lagrangian function and quasi-Newton updates of their Hessians. Since SQP methods are only locally convergent, some authors use hybrid techniques and combine robust globally convergent ascent methods with the SQP approach. For further details, the reader is referred to (Hettich and Kortanek, 1993) and references therein.

4.2.3. Adaptation of the exchange SIP method for the T-optimum design problem

Due to the relatively simple idea and, in consequence, the simplicity of the implementation, the algorithm proposed here to solve the problem (4.13) with (4.20) and (4.21) makes use of the exchange method. Since the use of global optimizers is necessary to ensure convergence, to this end the Adaptive Random Search (ARS) scheme has been integrated into the algorithm, cf. Section 4.4 for further details. The k-th auxiliary problem has the form

$$\mathcal{S}(\Theta_2^k) \begin{cases} \min\limits_{u \in U^k} f(u) \\ U^k = \Big\{ u \in U \in \mathbb{R}^{ns+n+1} : G(u, \vartheta_2) \leq 0, \ \forall \vartheta_2 \in \Theta_2^k \Big\}, \end{cases} \tag{4.38}$$

where u, f and G have the forms given by (4.20) and (4.21). Then the proposed algorithm is as follows:

Algorithm 4.2 (SIP based algorithm for T-optimum designs)**.**

During the k-th iteration:

Step 1: Given $\Theta_2^k \subset \Theta$, find a global solution $\widehat{u}^k$ of the problem $\mathcal{S}(\Theta_2^k)$.

Step 2: Find a global maximum $\widehat{\vartheta}_2^k$ of the subproblem $\mathcal{Q}(\widehat{u}^k)$ defined as follows:

$$\mathcal{Q}(\widehat{u}^k): \quad \max_{\vartheta_2 \in \Theta_2} G(\widehat{u}^k, \vartheta_2). \tag{4.39}$$

Step 3: If

$$G(\widehat{u}^k, \widehat{\vartheta}_2^k) \leq 0, \tag{4.40}$$

then STOP ($\widehat{u}^k$ is the optimum solution), otherwise construct Θ_2^{k+1} such that

$$\Theta_2^{k+1} \subset \Theta_2^k \cup \{\widehat{\vartheta}_2^k\}, \tag{4.41}$$

set $k \leftarrow k+1$ and go to Step 1 (repeat the procedure again).

The strategy of constructing the set Θ_2^k consists in simply adding a single most violated constraint and no deletion of the existing constraints. Since usually the set Θ_2, defining the admissible parameter values of the alternative model, is a hypercube, the simplest way to construct the starting set Θ_2^1 is to form it from the centre of this hypercube.

If we only assume that there exists $0 < \Gamma < \infty$ such that

$$U^k(\Theta_2^k) = \Big\{ u \in U : G(u, \vartheta_2) \leq 0, \ \forall \vartheta_2 \in \Theta_2^k \text{ and } \|u\|_\infty \leq \Gamma \Big\}, \tag{4.42}$$

i.e., the solution of $\mathcal{Q}(\widehat{u}^k)$ is bounded and U^k is a compact set contained in a ball around zero with radius Γ, the convergence of the proposed scheme can be proved in much the same way as in (Hettich and Kortanek, 1993, pp. 410–411). Note that in our case this condition is satisfied.

During the first step of the algorithm, the ARS scheme is used as the global optimizer for the constrained problem $\mathcal{S}(\Theta_2^k)$. To this end the conversion of $\mathcal{S}(\Theta_2^k)$ to an equivalent unconstrained problem is performed. We take advantage of the fact that the original problem

$$\begin{cases} \min\limits_{\xi\in\Xi(X),\beta} \beta \\ \text{subject to} \quad -\beta - \mathcal{J}(\xi,\vartheta_2) \le 0, \quad \forall \vartheta_2 \in \Theta_2^k, \end{cases} \tag{4.43}$$

is equivalent to

$$\beta^\star = \min_{\xi\in\Xi(X)} \left\{ - \min_{\vartheta_2\in\Theta_2^k} \Big(\mathcal{J}(\xi,\vartheta_2) \Big) \right\}. \tag{4.44}$$

Since Θ_2 is most often a hypercube, the ARS can be directly used as the global optimizer during the search for the most violated constraint in the second step of the algorithm.

When dealing with multiresponse dynamic systems models, we must remember that the most costly part of the target function evaluation during optimization is concentrated on numerical calculation of the responses of the alternative model for varying parameter values. Thus for the efficiency of an implementation it is crucial to avoid such response calculation as much as possible. Hence, during the optimization performed in the second step of the proposed scheme, the responses η_2 should be computed for each $\vartheta_2 \in \Theta_2^k$ once and stored in computer memory. Unfortunately, optimization performed during the second step of algorithm with respect to the parameters of the alternative model requires such calculations each time once the target function is evaluated (a continuous search over Θ_2 is unavoidable).

Remark 6. Generally, bear in mind that, e.g., if we deal with the models of distributed parameter systems described by adequate partial differential equations (PDEs), the evaluation of the target function is involved with a sequence of unavoidable numerical approximations (numerical solution of the PDE, its interpolation between grid points, integration of the complex criterion function, etc.). The accuracy of such approximations affects the optimization process and, consequently, implicitly the accuracy of the overall solution (optimum design). However, it is obvious that an improvement in accuracy can drastically increase computation time. Fortunately, by applying the Equivalence Theorem, we are always able to check the quality of the obtained solutions.

The above comment remains valid in the case of any algorithm used in the numerical search for approximations of optimum designs.

Example 4.3. Consider the state-space model of a simple pendulum

$$\begin{aligned} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} &= \omega(t), & \varphi(0) &= \pi/4, \\ \frac{\mathrm{d}\omega(t)}{\mathrm{d}t} &= -10\sin(\varphi(t)), & \omega(0) &= 0, \end{aligned} \tag{4.45}$$

where $\varphi(t)$ stands for the angle of pendulum swing at time t and $\omega(t)$ denotes the angular velocity. The alternative model has the linear structure

$$\begin{aligned} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} &= \omega(t), & \varphi(0) &= \pi/4, \\ \frac{\mathrm{d}\omega(t)}{\mathrm{d}t} &= -\vartheta_2\varphi(t) & \omega(0) &= 0, \end{aligned} \tag{4.46}$$

where ϑ_2 stands for an unknown, but constant, parameter value. It was assumed that the design region was $T = [0, 0.94]$, the set of admissible parameter values $\Theta_2 \in [7, 11]$, and the design size $n = 3$ (i.e., it consisted of three time instants). Using the approach described above, we obtained the optimum design

$$\xi^\star = \begin{Bmatrix} 0.1789, & 0.5309, & 0.1789 \\ 0.2034, & 0.5351, & 0.2615 \end{Bmatrix}, \tag{4.47}$$

which obviously reduces to the two-point design

$$\xi^\star = \begin{Bmatrix} 0.1789, & 0.5309 \\ 0.4649, & 0.5351 \end{Bmatrix}. \tag{4.48}$$

The worst value of the alternative model parameter with respect to $\xi^\star$ is $\vartheta_2^\star = 9.2957$. Figure 4.3 shows the respective sensitivity function $\psi(t, \xi^\star) = \|\eta(t) - \eta_2(t, \vartheta_2^\star)\|^2$ as a function of time and the responses of both the models. The program to calculate this numerical approximation of the optimum design was written using Lahey-Fujitsu Fortran 95 compiler v.5.6 with the aid of procedures from IMSL Fortran 90 MP Library v.4.0 (e.g., the `DNCONF` function implementing the sequential quadratic programming (SQP) method for nonlinearly constrained optimization).

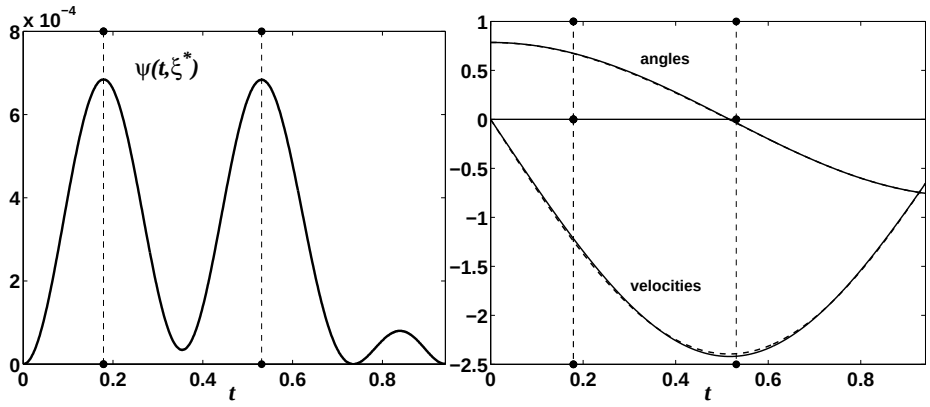
4.3. Relaxation procedure for solving T-optimum design problems

As was already indicated in (Fedorov and Hackl, 1997, p. 95), the classical sequential Wynn-Fedorov algorithm tailored to tackle the T-optimality criterion may suffer from the lack of convergence if a global minimizer for the problem

$$\min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x), \tag{4.49}$$

which has to be solved when computing $\Delta_1(\xi)$, is not unique. Although this phenomenon has been well known for many years, no viable alternative for this scheme has been proposed in the optimum experimental design literature so far. The only improvements concern a kind of regularization recommended in (Fedorov and Hackl, 1997). Namely, the authors suggest to replace successive designs ξ^k by $\bar{\xi}^k = (1 - \gamma)\xi^k + \gamma\bar{\xi}$, where $0 < \gamma \ll 1$ and $\bar{\xi}$ is a regular design, i.e., the design for which the minimization problem

$$\min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \bar{\xi}(\mathrm{d}x), \tag{4.50}$$



(a) (b)

Fig. 4.3. Sensitivity function (a) and true and alternative model responses (continuous and dashed lines, respectively) (b) with the location of the support points (dashed vertical lines) obtained in Example 4.3. The upper two lines in (b) stand for the angle φ, whereas the lower ones signify the angular velocity ω.

possesses a unique solution. But this is only a vague hint whose utility has never been formally proved and it can only be qualified as a heuristic which may fail (at any rate, it is possible to provide examples when this is the case).

In what follows, we outline a relaxation procedure which is robust to the presence of non-unique minimizers for (4.49). What is more, we give the proof of its convergence in a finite number of steps. A similar procedure has been used for numerically solving semi-infinite programming problems since the 1970s, cf. (Shimizu and Aiyoshi, 1980), but in the context of Euclidean spaces. Conceptually, it is an exchange SIP method as discussed in Section 4.2.2. The most serious complication which seems to prevent its use for determination of T-optimum designs is the necessity of operating on a space of probability measures in lieu of a finite-dimensional Hilbert space. Yet, below, we show that this obstacle can be overcome at the cost of additional efforts put forth to manage the technicalities related to the notion of weakly convergent probability measures. As a result, the original T-optimum design problem with a constraint set Θ_2 consisting of an infinite number of elements is decomposed into a sequence of substantially simpler 'relaxed' T-optimum design problems with finite-element constraint sets. The presented technique can be regarded as a kind of nonlinear cutting plane method (Bertsekas, 1999).

4.3.1. Relaxation algorithm and its convergence

Algorithm RATO (Relaxation Algorithm for T-Optimality)

Algorithm 4.3 (RATO)**.**

Step 1: (Initialization)
Choose any initial point $\vartheta_2^1 \in \Theta_2$ and define the initial set of representative parameter values $\Theta_2^1 = \{\vartheta_2^1\}$. Select $0 < \epsilon \ll 1$, a parameter used in the stopping rule, and set $k = 1$.

Step 2: (Solution of the relaxed problem)
Determine

$$\xi^k = \arg \max_{\xi \in \Xi(X)} \Delta_1^k(\xi), \tag{4.51}$$

where

$$\Delta_1^k(\xi) = \min_{\vartheta_2 \in \Theta_2^k} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x). \tag{4.52}$$

Step 3: (Determination of a candidate for the next representative parameter)
Find

$$\vartheta_2^{k+1} = \arg \min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi^k(\mathrm{d}x) \tag{4.53}$$

and store the corresponding value

$$\Delta_1(\xi^k) = \int_X \|\eta(x) - \eta_2(x, \vartheta_2^{k+1})\|^2 \, \xi^k(\mathrm{d}x). \tag{4.54}$$

Step 4: (Termination check)
Terminate if $\Delta_1(\xi^k) \geq \Delta_1^k(\xi^k) - \epsilon$. Otherwise, set $\Theta_2^{k+1} = \Theta_2^k \cup \{\vartheta_2^{k+1}\}$, replace k by $k + 1$ and go back to Step 2.

Roughly speaking, the algorithm is initiated from $\Theta_2^1 = \{\vartheta_2^1\}$ and then it iteratively seeks to maximize $\Delta_1(\,\cdot\,)$ by alternating between two phases: maximization of the 'relaxed' T-optimality criterion, i.e., the criterion computed for a finite set of representative parameters Θ_2^k, and then comparison of the so obtained optimal value with that of the genuine T-optimality criterion at the same 'relaxed' T-optimum design. If both the values do not differ (up to a small positive scalar ϵ standing for the desired degree of accuracy), then we terminate the iterations. Otherwise, the minimizer determining the value of the genuine T-optimality criterion at the current 'relaxed' T-optimum design is added to the set of representative parameters and the process of maximization is then rerun. This procedure of maximization and verification is thus repeated over and over.

It is a simple matter to show that the sequence of 'relaxed' values $\{\Delta_1^k(\xi^k)\}$ is nonincreasing, i.e., $\Delta_1^{k+1}(\xi^{k+1}) \leq \Delta_1^k(\xi^k)$ for each $k = 1, 2, \ldots$ If the above computational scheme is convergent, then $\Delta_1^k(\xi^k) \to \Delta_1(\xi^\star)$ from above, where $\xi^\star$ is a sought T-optimum design. In order to show this and to validate the termination criterion, observe first that, for each $\xi \in \Xi(X)$, we have that

$$\min_{\vartheta_2 \in \Theta_2^k} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x) \geq \min_{\vartheta_2 \in \Theta_2} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x), \tag{4.55}$$

which implies

$$\Delta_1^k(\xi) \geq \Delta_1(\xi), \quad \forall \xi \in \Xi(X). \tag{4.56}$$

Taking the maxima of both the sides, we get

$$\Delta_1^k(\xi^k) \geq \Delta_1(\xi^\star) \geq \Delta_1(\xi^k), \tag{4.57}$$

the latter inequality resulting from the definition of $\xi^\star$ as a global maximizer of $\Delta_1(\,\cdot\,)$. Consequently, if it happens that $\Delta_1^k(\xi^k) \approx \Delta_1(\xi^k)$, then we can rightly claim that we are reasonably close to $\xi^\star$ in the sense that $\Delta_1(\xi^k) \approx \Delta_1(\xi^\star)$.

Theorem 4.1. *Let Assumptions (A1)–(A3) of p. 41 hold. Then, for any given $\epsilon > 0$, Algorithm RATO terminates in a finite number of iterations.*

Proof. Denote

$$\mathcal{J}(\xi, \vartheta_2) = \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi(\mathrm{d}x). \tag{4.58}$$

Since $\Theta_2^k \subset \Theta_2^{k+1}$, we have

$$\min_{\vartheta_2 \in \Theta_2^k} \mathcal{J}(\xi, \vartheta_2) \geq \min_{\vartheta_2 \in \Theta_2^{k+1}} \mathcal{J}(\xi, \vartheta_2), \quad \forall \xi \in \Xi(X), \tag{4.59}$$

or equivalently,

$$\Delta_1^k(\xi) \geq \Delta_1^{k+1}(\xi), \quad \forall \xi \in \Xi(X). \tag{4.60}$$

Taking the maxima of both the sides over $\Xi(X)$ yields

$$\Delta_1^k(\xi^k) \geq \Delta_1^{k+1}(\xi^{k+1}), \tag{4.61}$$

i.e., the sequence $\left\{\Delta_1^k(\xi^k)\right\}$ is nonincreasing and zero is its lower bound.

From this, the sequential weak compactness of $\Xi(X)$ and the compactness of Θ_2, there must be a subsequence $\left\{\xi^{k^i}\right\}$ of $\left\{\xi^k\right\}$ such that

$$\xi^{k^i} \xrightarrow[i\to\infty]{} \bar{\xi} \quad \text{weakly}, \tag{4.62}$$

$$\Delta_1^{k^i}(\xi^{k^i}) \xrightarrow[i\to\infty]{} \bar{\Delta}_1, \tag{4.63}$$

$$\vartheta_2^{k^i+1} \xrightarrow[i\to\infty]{} \bar{\vartheta}_2 \tag{4.64}$$

for some $\bar{\xi} \in \Xi(X)$, $\bar{\Delta}_1 \geq 0$ and $\bar{\vartheta}_2 \in \Theta_2$.

Consider the relaxed problem in the (k^{i+1})-th iteration. Clearly, since $k^{i+1} \geq k^i + 1$, we see that $\vartheta_2^{k^i+1} \in \Theta_2^{k^{i+1}}$ and hence

$$\mathcal{J}(\xi^{k^{i+1}}, \vartheta_2^{k^i+1}) \geq \Delta_1^{k^{i+1}}(\xi^{k^{i+1}}), \quad \forall i \in \mathbb{N}. \tag{4.65}$$

Letting $i \to \infty$ and applying Lemma 3.2, we obtain

$$\mathcal{J}(\bar{\xi}, \bar{\vartheta}_2) \geq \bar{\Delta}_1. \tag{4.66}$$

Observe that, for each $i = 1, 2, \ldots$, we have that

$$\vartheta_2^{k^i+1} \in \widehat{\Theta}_2(\xi^{k^i}) = \operatorname{Arg}\min_{\theta_2 \in \Theta_2} \mathcal{J}(\xi^{k^i}, \vartheta_2) \tag{4.67}$$

and hence $\bar{\vartheta}_2$ is a cluster point of the sequence $\left\{\widehat{\Theta}_2(\xi^{k^i})\right\}$. The sequential weak outer semicontinuity of $\widehat{\Theta}_2(\,\cdot\,)$, cf. Theorem 3.3, then implies

$$\bar{\vartheta}_2 \in \operatorname*{Lim\,sup}_{i\to\infty} \widehat{\Theta}_2(\xi^{k^i}) \subset \widehat{\Theta}_2(\bar{\xi}). \tag{4.68}$$

Consequently,

$$\Delta_1(\bar{\xi}) = \min_{\vartheta_2\in\Theta_2} \mathcal{J}(\bar{\xi},\theta_2) = \mathcal{J}(\bar{\xi},\bar{\vartheta}_2). \tag{4.69}$$

From (4.66) we thus have

$$\Delta_1(\bar{\xi}) \geq \bar{\Delta}_1. \tag{4.70}$$

The weak continuity of $\Delta_1(\,\cdot\,)$ yields

$$\lim_{i\to\infty} \Delta_1(\xi^{k^i}) = \Delta_1(\bar{\xi}) \geq \bar{\Delta}_1. \tag{4.71}$$

This means that there exists some index i' such that

$$\Delta_1(\xi^{k^i}) \geq \bar{\Delta}_1 - \frac{\epsilon}{2}, \quad \forall i \geq i'. \tag{4.72}$$

On the other hand, from (4.63) we deduce that there must be some index i'' for which

$$\bar{\Delta}_1 \geq \Delta_1^{k^i}(\xi^{k^i}) - \frac{\epsilon}{2}, \quad \forall i \geq i''. \tag{4.73}$$

Therefore, for any $i \geq \max\{i', i''\}$, we have

$$\Delta_1(\xi^{k^i}) \geq \bar{\Delta}_1 - \frac{\epsilon}{2} \geq \Delta_1^{k^i}(\xi^{k^i}) - \frac{\epsilon}{2} - \frac{\epsilon}{2} = \Delta_1^{k^i}(\xi^{k^i}) - \epsilon, \tag{4.74}$$

i.e., the termination criterion in Step 4 is satisfied in a finite number of iterations. This completes the proof. □

It goes without saying that, in order to implement Algorithm RATO, it is necessary to detail numerical solution of the relaxed problem given in Step 2.

4.3.2. Numerical maximization of the 'relaxed' T-optimum design criterion

The following result turns out to be extremely useful while numerically solving the 'relaxed' T-optimum design problem

Theorem 4.2. *Under Assumptions (A1)–(A3), a purely discrete design*

$$\xi^k = \arg \max_{\xi\in\Xi(X)} \Delta_1^k(\xi) \tag{4.75}$$

exists, comprising no more than $k+1$ support points.

Proof. Fixing $\varsigma = \Delta_1^k(\xi^k)$, we have that the design ξ^k satisfies the following system of linear inequalities:

$$\begin{cases} \int_X f_1(x)\,\xi^k(\mathrm{d}x) \geq \varsigma, \\ \qquad\qquad \vdots \\ \int_X f_k(x)\,\xi^k(\mathrm{d}x) \geq \varsigma, \end{cases} \tag{4.76}$$

where $f_i(x) = \|\eta(x) - \eta_2(x, \vartheta_2^i)\|^2$, which can equivalently be represented in vector form

$$\int g(x)\,\xi^k(\mathrm{d}x) \geq 0 \tag{4.77}$$

with $g(x) = (f_1(x) - \varsigma, \ldots, f_k(x) - \varsigma)$.

Then the left-hand side of (4.77) defines an element of the convex hull of the set $\mathcal{S} = \{g(x) : x \in X\} \subset \mathbb{R}^k$, defined as

$$\mathrm{conv}(\mathcal{S}) = \left\{ \int_X g(x)\,\xi(\mathrm{d}x) : \xi \in \Xi(X) \right\}. \tag{4.78}$$

Hence, from Carathéodory's theorem (Silvey, 1980, p.72), $\int_X g(x)\,\xi^k(\mathrm{d}x)$ can be expressed as a convex combination of no more than $k+1$ elements of $\mathcal{S}$, i.e.,

$$\int_X g(x)\,\xi^k(\mathrm{d}x) = \sum_{i=1}^{k+1} p_i^k g(x_i^k), \tag{4.79}$$

where

$$p_i^k \geq 0, \quad i = 1, \ldots, k+1, \quad \sum_{i=1}^{k+1} p_i^k = 1, \tag{4.80}$$

x_i^k, $i = 1, \ldots, k+1$ being some points of X. This means that we may rightly identify ξ^k with the approximate discrete design

$$\begin{Bmatrix} x_1^k, & \ldots, & x_{k+1}^k \\ p_1^k, & \ldots, & p_{k+1}^k \end{Bmatrix}. \tag{4.81}$$

□

The above theorem simplifies the maximization of Step 2 to a search among discrete designs with a limited number of support points. This can be cast as the finite-dimensional maximin problem

$$\max_{x_i, p_i} \min \left\{ \sum_{i=1}^{k+1} p_i \|\eta(x_i) - \eta_2(x_i, \vartheta_2^1)\|^2, \ldots, \sum_{i=1}^{k+1} p_i \|\eta(x_i) - \eta_2(x_i, \vartheta_2^k)\|^2 \right\}, \tag{4.82}$$

subject to

$$\begin{aligned} x_i &\in X, \quad i = 1, \dots, k+1, \\ p_i &\geq 0, \quad i = 1, \dots, k+1, \\ \sum_{i=1}^{k+1} p_i &= 1, \end{aligned} \tag{4.83}$$

or the equivalent smooth nonlinear programming problem

$$\max_{x_i, p_i, \varsigma} \varsigma \tag{4.84}$$

subject to

$$\begin{aligned} &\sum_{i=1}^{k+1} p_i \|\eta(x_i) - \eta_2(x_i, \vartheta_2^\ell)\|^2 \geq \varsigma, \quad \ell = 1, \dots, k, \\ &x_i \in X, && i = 1, \dots, k+1, \\ &p_i \geq 0, && i = 1, \dots, k+1, \\ &\sum_{i=1}^{k+1} p_i = 1. \end{aligned} \tag{4.85}$$

The optimality of ξ^k can be checked using the general optimality conditions given in Theorem 3.7. Since Θ_2^k has only k elements, probability measures ζ on

$$\widehat{\Theta}_2^k(\xi^k) = \operatorname{Arg} \min_{\vartheta_2 \in \Theta_2^k} \int_X \|\eta(x) - \eta_2(x, \vartheta_2)\|^2 \, \xi^k(\mathrm{d}x) = \{\vartheta_2^{i_1}, \dots, \vartheta_2^{i_\ell}\}$$

are just discrete and (3.64) converts into the following set of conditions:

$$\min_{q_1, \dots, q_\ell} \max_{x \in X} \sum_{j=1}^{\ell} q_j \|\eta(x) - \eta_2(x, \vartheta_2^{i_j})\|^2 \leq \Delta_1^k(\xi^k) \tag{4.86}$$

subject to

$$\begin{aligned} q_j &\geq 0, \quad j = 1, \dots, \ell, \\ \sum_{j=1}^{\ell} q_j &= 1. \end{aligned} \tag{4.87}$$

Their verification may constitute a rather costly problem from the computational viewpoint if k gets large and X is high-dimensional. Therefore, below we present a simple approach which consists in smoothing the 'relaxed' T-optimality criterion.

4.3.2.1. Smooth concave approximation to the 'min' function

It is well known that the function $f : \mathbb{R}^k \to \mathbb{R}$ defined as

$$f(v) = \min\{v_1, \ldots, v_k\} \tag{4.88}$$

is not differentiable, which inevitably gives rise to extreme difficulties in extending classical optimization methods often using the information of the gradient and Hessian of the objective function.

In what follows, we propose to approximate it using the following parameterized, smooth exponential penalty function:

$$f_\lambda(v) = -\lambda \ln\left(\sum_{i=1}^{k} \exp\left(-\frac{v_i}{\lambda}\right)\right), \tag{4.89}$$

where $\lambda > 0$ is a parameter.

An analogous function was originally proposed as an exponential penalty function for constrained optimization, and then it was successfully applied in a smoothing method for minimax problems (Polak, 1997; Chen *et al.*, 2004; Xu, 2001). The following result characterizes its interesting uniform approximation properties.

Lemma 4.1. *We have the following properties:*

(i) Given $v \in \mathbb{R}^k$, the mapping $\lambda \mapsto f_\lambda(v)$ is nonincreasing in $(0, \infty)$.

(ii) For all $v \in \mathbb{R}^k$ we have

$$f(v) \geq f_\lambda(v) \geq f(v) - \lambda \ln(k). \tag{4.90}$$

(iii) Given $\lambda > 0$, $f_\lambda(\,\cdot\,)$ is twice continuously differentiable. Its gradient and Hessian are

$$\nabla f_\lambda(v) = \mu_\lambda(v), \tag{4.91}$$

$$\nabla^2 f_\lambda(v) = \frac{1}{\lambda}\left[\mu_\lambda(v)\mu_\lambda^{\mathrm{T}}(v) - \mathrm{Diag}(\mu_\lambda(v))\right], \tag{4.92}$$

respectively, where

$$\mu_\lambda(v) = \mathrm{col}[\mu_{\lambda 1}(v), \ldots, \mu_{\lambda k}(v)], \tag{4.93}$$

$$\mu_{\lambda i}(v) = \frac{\exp(-v_i/\lambda)}{\sum_{j=1}^{k} \exp(-v_j/\lambda)} \in (0,1), \quad i = 1, \ldots, k, \tag{4.94}$$

and $\mathrm{Diag}(\mu_\lambda(v))$ *signifies the diagonal matrix with its i-th diagonal entry $\mu_{\lambda i}(v)$ for each $i = 1, \ldots, k$.*

(iv) Given $\lambda > 0$, $f_\lambda(\,\cdot\,)$ is concave.

(v) Given $\lambda > 0$, $f_\lambda(\,\cdot\,)$ is nondecreasing, i.e., $f_\lambda(v) \leq f_\lambda(w)$ whenever $v \leq w$.

Proof. Part (i): Given $v \in \mathbb{R}^k$, it may be concluded that

$$\begin{aligned}\frac{\partial}{\partial\lambda} f_\lambda(v) &= \frac{\partial}{\partial\lambda}\left[-\lambda \ln\left(\sum_{i=1}^{k} \exp\left(-\frac{v_i}{\lambda}\right)\right)\right] \\ &= -\ln\left(\sum_{j=1}^{k} \exp\left(-\frac{v_j}{\lambda}\right)\right) - \sum_{i=1}^{k} \frac{v_i}{\lambda}\mu_{\lambda i}(v),\end{aligned} \tag{4.95}$$

where the weighting coefficients $\mu_{\lambda i}(v)$ are defined by (4.94). Since the $\mu_{\lambda i}$'s sum up to unity, it follows that

$$\frac{\partial}{\partial\lambda} f_\lambda(v) = -\sum_{i=1}^{k}\left[\frac{v_i}{\lambda} + \ln\left(\sum_{j=1}^{k} \exp\left(-\frac{v_j}{\lambda}\right)\right)\right]\mu_{\lambda i}(v) \le 0 \tag{4.96}$$

because

$$\ln\left(\sum_{j=1}^{k} \exp\left(-\frac{v_j}{\lambda}\right)\right) \ge \ln\left(\exp\left(-\frac{v_i}{\lambda}\right)\right) = -\frac{v_i}{\lambda}, \quad i = 1, \ldots, k. \tag{4.97}$$

Consequently, $\lambda \mapsto f_\lambda(v)$ is nonincreasing.

Part (ii): By definition, given $\lambda > 0$, we have

$$\begin{aligned}f(v) - f_\lambda(v) &= \lambda \ln\left(\exp\left(\frac{f(v)}{\lambda}\right)\right) + \lambda \ln\left(\sum_{i=1}^{k} \exp\left(-\frac{v_i}{\lambda}\right)\right) \\ &= \lambda \ln\left(\sum_{i=1}^{k} \exp\left(\frac{\min\{v_1, \ldots, v_k\} - v_i}{\lambda}\right)\right).\end{aligned} \tag{4.98}$$

But

$$1 \le \sum_{i=1}^{k} \exp\left(\frac{\min\{v_1, \ldots, v_k\} - v_i}{\lambda}\right) \le k, \tag{4.99}$$

which gives

$$0 \le f(v) - f_\lambda(v) \le \lambda \ln(k). \tag{4.100}$$

Part (iii): Formulae (4.91) and (4.92) are easy to verify by direct differentiation.

Part (iv): The assertion will be shown if we prove that the Hessian $\nabla^2 f_\lambda(v)$ is negative semidefinite for all $v \in \mathbb{R}^k$, i.e., if

$$y^{\mathrm{T}} \nabla^2 f_\lambda(v) y \le 0, \quad \forall y \in \mathbb{R}^k. \tag{4.101}$$

From Part (iii) we deduce that

$$y^{\mathrm{T}} \nabla^2 f_\lambda(v) y = \frac{1}{\lambda}\left[\left(\sum_{i=1}^{k} \alpha_i y_i\right)^2 - \sum_{i=1}^{k} \alpha_i y_i^2\right], \tag{4.102}$$

where $\alpha_i = \mu_{\lambda i}(v)$, $i = 1, \ldots, k$. Observing that $\alpha_1, \ldots, \alpha_k > 0$ and $\sum_{i=1}^k \alpha_i = 1$, for any convex function $g : \mathbb{R} \to \mathbb{R}$ we must have

$$g\left(\sum_{i=1}^k \alpha_i y_i\right) \le \sum_{i=1}^k \alpha_i g(y_i), \tag{4.103}$$

which is a special case of Jensen's inequality (Pollard, 2002, p. 29). Setting $g(z) = z^2$, we get

$$\left(\sum_{i=1}^k \alpha_i y_i\right)^2 - \sum_{i=1}^k \alpha_i y_i^2 \le 0, \tag{4.104}$$

which implies (4.101).

Part (v): Let $v, w \in \mathbb{R}^k$ satisfy $v \le w$ and $\lambda > 0$ be fixed. Then for each $i = 1, \ldots, k$ we have $v_i \le w_i$ and thereby $-v_i/\lambda \ge -w_i/\lambda$. Since both the exponential and logarithm functions are increasing, it follows that

$$\lambda \ln\left(\sum_{i=1}^k \exp\left(-\frac{v_i}{\lambda}\right)\right) \ge \lambda \ln\left(\sum_{i=1}^k \exp\left(-\frac{w_i}{\lambda}\right)\right), \tag{4.105}$$

which yields the desired conclusion after multiplying both the sides by -1. □

4.3.2.2. Smoothing method for maximization of $\Delta_1^k(\,\cdot\,)$

Note that writing

$$\mathcal{J}_i(\xi) = \int_X \|\eta(x) - \eta_2(x, \vartheta_2^i)\|^2 \, \xi(\mathrm{d}x), \quad i = 1, \ldots, k, \tag{4.106}$$

we obtain

$$\Delta_1^k(\xi) = \min\left\{\mathcal{J}_1(\xi), \ldots, \mathcal{J}_k(\xi)\right\}. \tag{4.107}$$

Clearly, this form coincides with (4.88) and this observation, owing to excellent approximation properties of the exponential penalty function (4.89), cf. Part (ii) of Lemma 4.1, suggests transformation of the T-optimum design problem into the smooth problem consisting in maximizing

$$\Delta_{1\lambda}^k(\xi) = -\lambda \ln\left(\sum_{i=1}^k \exp\left(-\frac{\mathcal{J}_i(\xi)}{\lambda}\right)\right) \tag{4.108}$$

for a fixed $0 < \lambda \ll 1$. The probability measures maximizing $\Delta_{1\lambda}^k(\,\cdot\,)$ will be called T_λ^k-optimum designs.

The essential fact is that the design criterion $\Delta_{1\lambda}^k(\,\cdot\,)$ possesses the same properties as other smooth criteria commonly used in optimum experimental design. In particular, given $\xi \in \Xi(X)$, we can define the following vector being a counterpart of the Fisher information matrix:

$$v(\xi) = \mathrm{col}\big[\mathcal{J}_1(\xi), \ldots, \mathcal{J}_k(\xi)\big] = \int_X \varphi(x) \, \xi(\mathrm{d}x), \tag{4.109}$$

where

$$\varphi(x) = \operatorname{col}\big[\|\eta(x) - \eta_2(x, \vartheta_2^1)\|^2, \ldots, \|\eta(x) - \eta_2(x, \vartheta_2^k)\|^2\big], \tag{4.110}$$

and the set

$$\mathfrak{v}(X) = \{v(\xi) : \xi \in \Xi(X)\}. \tag{4.111}$$

Consequently, in much the same way as Lemmas 3.2 and 3.3 in (Uciński, 2005), we can prove the following two results.

Lemma 4.2. *The set $\mathfrak{v}(X)$ is compact and convex.*

Lemma 4.3 (Bound on the number of support points)**.** *For any $v_0 \in \mathfrak{v}(X)$ there always exists a purely discrete design ξ with no more than $k+1$ support points such that $v(\xi) = v_0$. If v_0 lies on the boundary of $\mathfrak{v}(X)$, then the number of support points is less than or equal to k.*

Lemma 4.3 justifies restricting our attention only to approximate discrete designs with a limited number of supporting points, and this proves again that the introduction of continuous designs, which may have seemed at first sight a superfluous complication, leads to implementable results from a practical point of view.

The following characterizations are of paramount importance:

Lemma 4.4. *Let Assumptions (A1)–(A3) hold. Given $\lambda > 0$, we have the following properties:*

(i) $\Delta_{1\lambda}^k(\,\cdot\,)$ *is concave.*

(ii) If $v, w \in \mathfrak{v}(X)$ *are such that* $v \le w$*, then* $\Delta_{1\lambda}^k(v) \le \Delta_{1\lambda}^k(w)$*. (Monotonicity)*

(iii) For any $\xi, \nu \in \Xi(X)$*, we have*

$$\frac{\partial}{\partial\lambda}\Delta_{1\lambda}^k[v((1-\alpha)\xi + \alpha\nu)]\bigg|_{\alpha=0^+} = \sum_{i=1}^k \mu_{\lambda i}(v(\xi))[\mathcal{J}_i(\nu) - \mathcal{J}_i(\xi)]. \tag{4.112}$$

Proof. Part (i): Since the mapping $f_\lambda(\,\cdot\,)$, cf. (4.89), is concave and the mapping $v(\,\cdot\,)$ is linear, their composite $f_\lambda(v(\,\cdot\,)) = \Delta_{1\lambda}^k(\,\cdot\,)$ must be concave.

Part (ii): It follows immediately from Part (v) of Lemma 4.1.

Part (iii): Part (iii) of Lemma 4.1, when combined with application of the chain rule of differentiation, yields

$$\begin{aligned}\frac{\partial}{\partial\lambda}\Delta_{1\lambda}^k[v((1-\alpha)\xi + \alpha\nu)] &= \frac{\partial}{\partial\lambda}\Delta_{1\lambda}^k[(1-\alpha)v(\xi) + \alpha v(\nu)] \\ &= \mu_\lambda^{\mathrm{T}}((1-\alpha)v(\xi) + \alpha v(\nu))(v(\nu) - v(\xi)) \\ &= \sum_{i=1}^k \mu_{\lambda i}((1-\alpha)v(\xi) + \alpha v(\nu))[\mathcal{J}_i(\nu) - \mathcal{J}_i(\xi)],\end{aligned} \tag{4.113}$$

which implies the postulated form of the directional derivative. □

The above properties make it legitimate to treat the T^k_λ-optimum design problem in the usual framework of the optimum experimental design theory for smooth optimality criteria (Atkinson and Donev, 1992; Fedorov and Hackl, 1997; Walter and Pronzato, 1997; Pukelsheim, 1993; Pázman, 1986; Uciński, 2005). Consequently, the following results can be directly obtained. Their proofs are standard, cf. e.g. (Uciński, 2005), and therefore they are omitted.

Theorem 4.3 (Equivalence theorem for $\boldsymbol{\Delta^k_\lambda(\,\cdot\,)}$)**.** *The following characterizations are equivalent in the sense that each implies the other two:*

(i) the design $\xi^\star$ *maximizes* $\Delta^k_\lambda(\,\cdot\,)$*,*

(ii) the design $\xi^\star$ *minimizes* $\max\limits_{x\in X}\sum\limits_{i=1}^{k}\mu_{\lambda i}(\xi)\big[\|\eta(x)-\eta_2(x,\vartheta_2^i)\|^2-\mathcal{J}_i(\xi)\big]$*, and*

(iii) $\sum\limits_{i=1}^{k}\mu_{\lambda i}(\xi^\star)\|\eta(x)-\eta_2(x,\vartheta_2^i)\|^2\le\sum\limits_{i=1}^{k}\mu_{\lambda i}(\xi^\star)\mathcal{J}_i(\xi^\star),\quad\forall x\in X,$

where $\mu_{\lambda i}(\xi)=\exp(-\mathcal{J}_i(\xi)/\lambda)/\sum_{j=1}^{k}\exp(-\mathcal{J}_j(\xi)/\lambda)$*,* $i=1,\ldots,k$*.*

Theorem 4.4. *Let Assumptions (A1)–(A3) hold. Then:*

(i) A T^k_λ*-optimal design exists comprising no more than* k *points (i.e., one less than predicted by Lemma 4.3).*

(ii) The set of T^k_λ*-optimal designs is convex.*

(iii) For any purely discrete T^k_λ*-optimal design* $\xi^\star$*, the function* $\sum_{i=1}^{k}\mu_{\lambda i}(\xi^\star)\|\eta(x)-\eta_2(x,\vartheta_2^i)\|^2$ *attains its maximal value equal to* $\sum_{i=1}^{k}\mu_{\lambda i}(\xi^\star)\mathcal{J}_i(\xi^\star)$ *at all support points corresponding to nonzero weights.*

As we have repeatedly stated, the equivalence theorems in optimum experimental design constitute the characterization of an optimal measure $\xi^\star$ which solves the design problem. Theorem 4.3 is not an exception in this respect. But, additionally, we have to address the question of how to generate $\xi^\star$. As was already mentioned in Chapter 2, there are several iterative schemes for this task, each of the following genre:

Algorithm 4.4 (Successive maximization of $\boldsymbol{\Delta^k_{1\lambda}(\,\cdot\,)}$)**.**

Step 1: Let ξ^0 be a discrete starting design. Fix $0<\lambda\ll 1$ and choose some positive tolerance $\epsilon\ll 1$. Set $q=0$.

Step 2: Find

$$\widehat{x}^q=\arg\max_{x\in X}\psi(x,\xi^q),\tag{4.114}$$

where

$$\psi(x,\xi^q)=\sum_{i=1}^{k}\mu_{\lambda i}(\xi^q)\|\eta(x)-\eta_2(x,\vartheta_2^i)\|^2.\tag{4.115}$$

If $\psi(\widehat{x}^k, \xi^q) \leq \sum_{i=1}^{k} \mu_{\lambda i}(\xi^q)\mathcal{J}_i(\xi^q) + \epsilon$, then STOP.

Step 3: For an appropriate value of $0 < \alpha^q < 1$, update the design measure in accordance with the rule

$$\xi^{q+1} = (1 - \alpha^q)\xi^q + \alpha^q \delta(\widehat{x}^q), \tag{4.116}$$

where $\delta(\widehat{x}^q)$ places measure one at point $\widehat{x}^q$. Increment q by one and go to Step 2.

In the same way as for the classical first-order algorithms in common use for many years, the following choices for the sequence $\{\alpha^q\}$ yield the convergence of the above algorithm:

(i) Diminishing stepsize (Wynn's algorithm):

$$\lim_{q\to\infty} \alpha^q = 0, \quad \sum_{q=0}^{\infty} \alpha^q = \infty. \tag{4.117}$$

(ii) Limited maximization rule (Fedorov's algorithm):

$$\alpha^q = \arg \max_{0<\alpha<1} \Delta_{1\lambda}^k \left((1 - \alpha)\xi^q + \alpha\delta(\widehat{x}^q)\right). \tag{4.118}$$

Example 4.4. A motivation for this numerical example comes from industrial processes diagnosis, cf. Section 6.5. A majority of methods for fault detection use a system model to extract departures from a normal state and, in consequence, to generate an appropriate residuum signal (Korbicz *et al.*, 2004). If a true mathematical description of the system is complicated or an adequate model is hard to implement, T-optimum experimental designs can be helpful in finding a simpler alternative model. The system of two tanks shown schematically in Fig. 4.4 was considered as an example of a nonlinear, dynamic multiresponse system.

The system consists of two connected cylindrical tanks and a coil pipe fulfilling a delay unit. The nominal outflow Q_n is situated in the second tank. The liquid is supplied into the first tank by a pump with flow speed Q_1. Both tanks are equipped with sensors measuring liquid levels N_1 and N_2. The valves V_E, V_1, V_2, V_3, V_4 are switched manually. During the nominal work V_1 and V_4 are closed. By an appropriate combination of valve switchings, we can model faults, i.e., leakiness of the first tank. The above system can be described by a set of equations forming the so-called flow model (Heiming and Lunze, 1999):

$$\begin{aligned} \frac{\mathrm{d}N_1(t)}{\mathrm{d}t} &= \frac{1}{A_1}\left(Q_1(t) - K_1\sqrt{N_1(t)}\right), \\ \frac{\mathrm{d}N_2(t)}{\mathrm{d}t} &= \frac{1}{A_2}\left(K_1\sqrt{N_1(t)} - K_2\sqrt{N_2(t)}\right), \end{aligned} \tag{4.119}$$

where N_1 and N_2 stand for liquid levels in the first and second tanks, respectively. A_1 and A_2 can be interpreted as an averaged-by-height areas of a cut for each tank

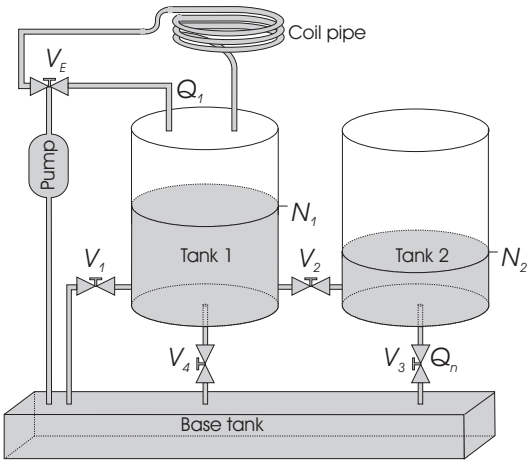


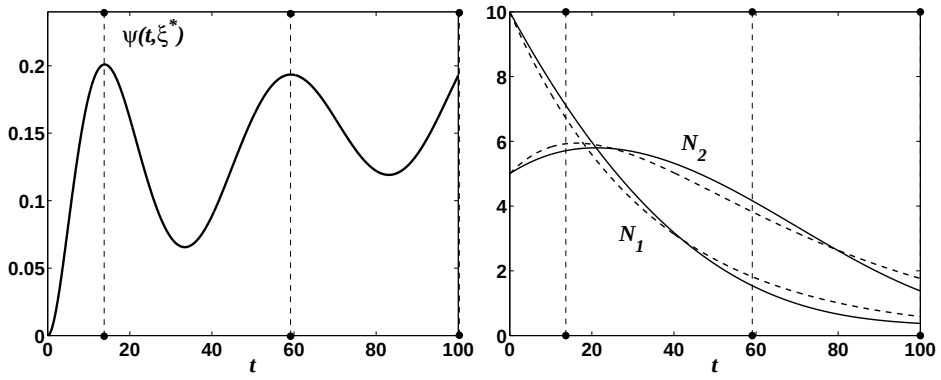
Fig. 4.4. Two tank system.

and parameters K_1 and K_2 as normalized averaged cuts of reducers supplying the liquid into tanks. The above model was considered as the true model of the two-tank system. The idea of the experiment is to obtain a design so that it allows for rejection of the alternative model, being in this case a linearized version of the flow model of the form

$$\begin{aligned}\frac{\mathrm{d}N_1(t)}{\mathrm{d}t} =& \frac{Q_1(t)}{\vartheta_{21}} - \frac{\vartheta_{23}}{\vartheta_{21}}\left(\sqrt{\vartheta_{25}} + \frac{N_1(t)-\vartheta_{25}}{2\sqrt{\vartheta_{25}}}\right),\\ \frac{\mathrm{d}N_2(t)}{\mathrm{d}t} =& \frac{\vartheta_{23}}{\vartheta_{22}}\left(\sqrt{\vartheta_{25}} + \frac{N_1(t)-\vartheta_{25}}{2\sqrt{\vartheta_{25}}}\right) - \frac{\vartheta_{24}}{\vartheta_{22}}\left(\sqrt{\vartheta_{25}} + \frac{N_2(t)-\vartheta_{25}}{2\sqrt{\vartheta_{25}}}\right).\end{aligned} \tag{4.120}$$

Parameters ϑ_{21} up to ϑ_{24} have the same interpretation as parameters A_1, A_2, K_1, K_2 of the true model. In both the models the influx Q_1 into the first tank can, in general, be time-varying. In the experiment this quantity was fixed at a constant level of $Q_1 = 15$ cm^3/s. The time horizon and initial values of the state variables were assumed as $t = [0, 100]$ and $N_1(0) = 10,\ N_2(0) = 5$, respectively. The parameters of the true model were fixed as $\vartheta_1 = (A_1,\ A_2,\ K_1,\ K_2) = (315.0,\ 315.0,\ 28.0,\ 28, 0)$, while the ranges for the parameter values in the alternative model were set as $\Theta_2 = \big\{[250, 350] \times [250, 350] \times [25, 35] \times [25, 35] \times [0.1, 100]\big\}$. The Fortran 95 program to calculate approximation of the optimum design using Algorithm 4.3 with $\epsilon = 0.01$ and Algorithm 4.4 incorporated in order to regularize the relaxed problems with $\lambda = 0.005$ was run using the Lahey-Fujitsu Fortran 95 compiler v5.6 with IMSL library. To solve the respective global optimization problem, the Adaptive Random Search (ARS) method was utilized. The resulting optimum design has the form

$$\xi^\star = \begin{Bmatrix} 13.67, & 59.08, & 100.00 \\ 0.434, & 0.251, & 0.315 \end{Bmatrix}. \tag{4.121}$$



(a) (b)

Fig. 4.5. Sensitivity function (a) and true and alternative model responses (b) (solid and dashed line, respectively) obtained in Example 4.4 (the location of the optimal support points is shown with dashed vertical lines).

This means that, e.g., about a quarter of the measuring effort should be assigned to the time $t = 59.08$ s. The corresponding least profitable vector of parameters of the alternative model was $\vartheta_2^\star = (325.51,\ 337.31,\ 27.47,\ 28.50,\ 0.48)$. The sensitivity function $\psi(t, \xi^\star) = \|\eta(t) - \eta_2(t, \vartheta_2^\star)\|^2$ has the form shown in Fig. 4.5. It is worth of noticing that the support points are located in the places where the sensitivity function achieves its upper bound equal to $\Delta_1(\xi^\star)$ which is consistent with the theory developed. The resultant design contains a minimal number of support points to guarantee non-singularity, i.e., the identifiability of the system parameters (the number of support points multiplied by the number of system outputs must be greater than or equal to the number of unknown system parameters in ϑ_2). The responses of the true and alternative models are also shown in Fig. 4.5.

4.4. Solving global optimization subtasks

A vital role in the iterative algorithms presented in this chapter is played by an efficient global optimization method. Since common nonlinear programming algorithms are known to converge to local optima, we have turned our attention to a stochastic optimization method called the adaptive random search (ARS), which is widely used in the engineering optimization literature (Walter and Pronzato, 1997). Based on numerous computer experiments it was found that this extremely simple strategy is especially suited for the purpose of global optimization problems arising while looking for numerical approximations of optimum designs. Thus it was used while implementing numerical procedures incorporated in the dissertation.

Originally, the algorithm solves a maximization problem $\max_{v \in V} J(v)$ for the admissible set V being a hypercube, i.e., the admissible range for v_i, $i = 1, \ldots, q$

is in the form

$$v_{i\,\min} \leq v_i \leq v_{i\,\max}. \tag{4.122}$$

We include the details from (Walter and Pronzato, 1997) in order to make our presentation self-contained.

The routine chooses the initial point v^0 at the centre of V. After r iterations, given the current best point v^r, a random displacement vector Δv is generated and the trial point

$$v^+ = \Pi_V(v^r + \Delta v) \tag{4.123}$$

is tested, where Δv follows a multinormal distribution with zero mean and covariance

$$\mathrm{Cov}\{\Delta v\} = \mathrm{diag}[\sigma_1, \ldots, \sigma_q], \tag{4.124}$$

Π_V being the projection onto V,

$$\Pi(v)_i = \begin{cases} v_{i\,\min} & \text{if } v_i < v_{i\,\min}, \\ v_i & \text{if } v_{i\,\min} \leq v_i \leq v_{i\,\max}, \\ v_{i\,\max} & \text{if } v_i > v_{i\,\max}. \end{cases} \tag{4.125}$$

If $J(v^+) < J(v^r)$ then v^+ is rejected and, consequently, we set $v^{r+1} = v^r$, otherwise v^+ is taken as v^{r+1}.

The adaptive strategy consists in repeatedly alternating two phases. During the first one (variance selection) the diagonal of $\mathrm{Cov}\{\Delta v\}$ is selected from among the sequence ${}^1\sigma, {}^2\sigma, \ldots, {}^5\sigma$, where

$${}^1\sigma = v_{\max} - v_{\min} \tag{4.126}$$

and

$${}^i\sigma = {}^{(i-1)}\sigma/10, \quad i = 2, \ldots, 5. \tag{4.127}$$

With such a choice, ${}^1\sigma$ is large enough to allow for an easy exploration of V, whereas ${}^5\sigma$ is small enough for a precise localization of an optimal point. In order to allow a comparison to be drawn, all the possible ${}^i\sigma$'s are used $100/i$ times, starting from the same initial value of v. The largest ${}^i\sigma$'s, designed to escape local maxima, are therefore used more often than the smaller ones.

During the second (exploration) phase, the most successful ${}^i\sigma$ in terms of the criterion value reached during the variance selection phase is used for 100 random trials started from the best v obtained so far. The variance-selection phase then resumes, unless the decision to stop is taken.

A detailed scheme of the ARS algorithm is as follows:

Algorithm 4.5 (Adaptive Random Search algorithm)**.**

Step 1: Initialization Choose v^0, N_{max}, i_{max}, j_{max} (usually $i_{max} = 5$, $j_{max} = 100$), ${}^1\sigma$, and set $v_{best} = v^0$, $N = 1$, $i = 1$.

Step 2: Variance-selection phase

2.1 Set $j = 1$, $v^j = v^0$ and ${}^i\sigma = {}^1\sigma 10^{-i+1}$.

2.2 Perturb v^j according to (4.123) to get a new trial point v^{j+}.

2.3 If $J(v^{j+}) \geq J(v^j)$ then $v^{j+1} = v^{j+}$, otherwise $v^{j+1} = v^j$.

2.4 If $J(v^{j+}) \geq J(v_{best})$ then $v_{best} = v^{j+}$, $i_{best} = i$.

2.5 If $j \leq j_{max}/i$ then set $j \leftarrow j + 1$ and go to **2.2**.

2.6 If $i < i_{max}$ then set $i \leftarrow i + 1$ and go to **2.1**.

Step 3: Variance-exploitation phase

3.1 Set $j = 1$, $v^j = v_{best}$, $i = i_{best}$ and ${}^i\sigma = {}^1\sigma 10^{-i+1}$.

3.2 Perturb v^j according to (4.123) to get a new trial point v^{j+}.

3.3 If $J(v^{j+}) \geq J(v^j)$ then $x^{j+1} = x^{j+}$, otherwise $v^{j+1} = v^j$.

3.4 If $J(v^{j+}) \geq J(v_{best})$ then $v_{best} = v^{j+}$.

3.5 If $j \leq j_{max}$ then set $j \leftarrow j + 1$ and go to **3.2**.

3.6 If $N = N_{max}$ then STOP.

3.6 Set $N \leftarrow N + 1$, $v^0 = v_{best}$ and repeat from **2.1**.

The parameter N_{max} allows for a possible multiple repetition of the whole 2-phase stage, but starting from the best point obtained so far.

The ARS does not use the information about the gradient of the performance index. Thus a significant numerical efficiency could hardly be expected. However, because of its valuable properties regarding global convergence and simplicity, the ARS seems to be more flexible and suitable in the case of dynamic systems (especially distributed parameter systems) than many classical non-linear programming approaches. Furthermore, gradient evaluation can be very costly or approximation of the gradient may fail to be satisfactory (e.g., there may occur some scaling problems or insufficient smoothness of the underlying functions). Nevertheless, the performance of the ARS can be improved by combination with various other methods. For example, since the smallest ${}^5\sigma$ corresponds to very small displacements in V, so occasionally we can switch to local maximization in order to make the results more accurate.

Two more issues require some comment. The first one concerns the usability of the ARS when solving constrained problems. As can be seen in Section 4.2, such problems can be transformed to unconstrained ones. Since the ARS belongs to the group of non-gradient methods, possible non-smoothness of the resulting transformed criterion function does not exclude the usability of the scheme. Another problem is that in practical situations the search space V may not be a hypercube (e.g., during the search over the spatial area Ω on which an adequate PDE describing some DPS is defined). That impediment can be overcome by embedding V in some set V' being a hypercube such that $V \subseteq V'$. Then a suitable penalty function can be constructed and the modified criterion can be maximized. Due to the reasons mentioned just before, the possible lack of smoothness of the resulting function to be maximized has no influence on the performance.

Finally, one more benefit of the ARS strategy is the possibility of its relatively simple parallelization, which is very important in the context of the efficiency of computations. We shall return to this question in Chapter 5

4.5. Concluding remarks

In summary, this chapter has dealt with three iterative numerical techniques for constructing approximations to T-optimum experimental designs. The key idea in all of them is to accomplish the solution by hill climbing on the T-optimality criterion. First, the classical Fedorov scheme tailored to the computation of T-optimal designs was discussed. Although it is practically the only tool to generate solutions in the literature involved with the considered design problem, it has been criticized to a significant measure even by its inventor. This is because the algorithm may diverge if the T-optimality criterion fails to be smooth, which is the case when, given a design ξ, the corresponding global minimizer in the set of admissible parameters for the alternative model, which is encountered in the definition of the T-optimality criterion, is not unique. A way out can be the conversion of the original problem into a semi-infinite programming one. This approach is very appealing in the light of the recent advances in this area of non-smooth optimization. Unfortunately, a weakness of this method is the lack of an adaptation mechanism regarding the number of support points, which should be set here *a priori*. Consequently, prior to computations, we must guess an upper bound on this number, which may result in excessive time of computations if the guess is too high, or in a nonoptimal solution if it is too small. The drawbacks of both the approaches were addressed in the final part of the chapter, where a novel relaxation algorithm, called RATO, was introduced that had been derived from an exchange method for solving SIP problems by a non-trivial generalization to the framework of the space of probability measures on X. The algorithm makes an iterative refinement to the existing solution at each step and adapts the number of support points if necessary. Since it reduces the problem to solving a sequence of simpler finite maximin problems, it is also easier to control the algorithm performance and to verify the optimality of the current design. The most distinctive feature is, however, the guaranteed convergence in a finite number of steps. In the author's opinion, this is the main contribution of this dissertation to the state of the art in algorithmic methods of optimum experimental design. Although the numerical examples presented to confirm the effectiveness of the presented approach are clearly not real-world problems and their purpose is primarily to illustrate our considerations in an easily interpretable manner, they are complex enough to provide evidence for the effectiveness of the proposed approach.

Chapter 5

SPECIAL TOPICS OF ALGORITHMIC OPTIMAL DESIGN

The T-optimum experimental design problem has been posed as an optimization one, that of maximizing a functional defined on a space of all admissible designs which are identified with probability measures on $\mathcal{B}(X)$, the sigma-algebra of Borel subsets of X, the design space. The indicated maximization is accomplished by the appropriate choice of a 'maximizing' sequence of designs, which can be obtained using the algorithms outlined in the previous chapter. It goes without saying that those methods constitute only a conceptual framework for prospective implementations in practice and numerous involved problems still have to be addressed. The purpose of this chapter is to indicate some of them for a variety of situations motivated by practical applications. Clearly, these considerations by no means cover all aspects of the reality and many difficult problems still remain open. Here we focus attention on settings which are expected to become fields of laborious research in the near future.

5.1. DT-optimum designs

Optimum designs for discrimination between models may have poor properties for estimation of parameters in chosen models. There exist a number of papers looking for a reasonable balance between model discrimination and parameter estimation (Biswas and Chaudhuri, 2002; Waterhouse *et al.*, 2004). However, there exists no ideal, universal solution to this problem. An interesting approach, called DT-optimality, combining properties of D- and T-optimum designs has recently been proposed by (Atkinson, 2005). It what follows, we briefly characterize it, propose a Wynn-Fedorov-type numerical algorithm for constructing DT-optimal designs and give a proof of its convergence.

To combine D- and T-optimality criteria, a common scale of comparison is necessary, since both the criteria are different in behaviour. To this end, Atkinson proposed the so-called efficiencies. The T-efficiency of any design ξ relative to T-optimum design $\xi^\star_{\mathrm{T}}$ is

$$E_f^{\mathrm{T}}(\xi) = \frac{\Delta_1(\xi)}{\Delta_1(\xi^\star_{\mathrm{T}})}, \tag{5.1}$$

whereas its D-efficiency relative to a D-optimum design $\xi_D^\star$ is

$$E_f^{\mathrm{D}}(\xi) = \left\{ \frac{\det(M_1(\xi))}{\det(M_1(\xi_{\mathrm{D}}^\star))} \right\}^{1/m_1}, \tag{5.2}$$

m_1 standing for the number of parameters in model $\mathcal{M}_1$ and $M_1(\xi)$ being the Fisher information matrix defined as

$$M_1(\xi) = \int_X f_1(x) f_1^{\mathrm{T}}(x)\, \xi(\mathrm{d}x) \tag{5.3}$$

and corresponding to model $\mathcal{M}_1$ linearized around the assumed 'true' parameter value $\widetilde{\vartheta}_1$, i.e., for every $x \in X$ there holds $\eta(x) = \eta_1(x, \widetilde{\vartheta}_1)$, so that we have

$$f_1(x) = \left(\frac{\partial \eta_1(x, \widetilde{\vartheta}_1)}{\partial \vartheta_1} \right)^{\mathrm{T}}. \tag{5.4}$$

Then, to obtain the appropriate optimality criterion for both discrimination and parameter estimation, Atkinson postulates to maximize a weighted product of the efficiencies:

$$\left(E_f^{\mathrm{T}}(\xi)\right)^{1-\kappa} \left(E_f^{\mathrm{D}}(\xi)\right)^{\kappa} = \left(\frac{\Delta_1(\xi)}{\Delta_1(\xi_{\mathrm{T}}^\star)} \right)^{1-\kappa} \left(\frac{\det(M_1(\xi))}{\det(M_1(\xi_{\mathrm{D}}^\star))} \right)^{\kappa/m_1}, \quad 0 \le \kappa \le 1, \tag{5.5}$$

where κ is a fixed coefficient weighting the importance attached to the particular criteria when combining them into one aggregate criterion. Note that for the extreme values of $\kappa = 0$ and $\kappa = 1$ we obtain T- and D-optimality, respectively. Then, taking logs in (5.5) and observing that terms involving $\xi_{\mathrm{T}}^\star$ and $\xi_{\mathrm{D}}^\star$ are constants when a maximum is found over $\xi(X)$, we obtain the following convenient form of the criterion to be maximized

$$\Phi_1^{\mathrm{DT}}(\xi) = (1-\kappa)\ln(\Delta_1(\xi)) + (\kappa/m_1)\ln\big(\det(M_1(\xi))\big), \tag{5.6}$$

and designs maximizing (5.6) are called the DT-optimum ones and are denoted by $\xi_{\mathrm{DT}}^\star$.

What is more, based on the fact that the DT-optimality criterion constitutes a linear combination of concave criteria, Atkinson proves the following equivalence theorem for the case of single-response models:

Theorem 5.1 (Equivalence theorem for DT-optimality)**.**

(i) A necessary and sufficient condition for a design $\xi_{\mathrm{DT}}^\star$ to be DT-optimum is the fulfilment of the inequality

$$\psi_1^{(\mathrm{DT})}(x, \xi_{\mathrm{DT}}^\star) \le 1, \quad x \in X, \tag{5.7}$$

where

$$\psi_1^{(\mathrm{DT})}(x, \xi) = \frac{1-\kappa}{\Delta_1(\xi)} \{\eta(x) - \eta_2(x, \vartheta_2)\}^2 + \frac{\kappa}{m_1} f_1^{\mathrm{T}}(x) M_1^{-1}(\xi) f_1(x). \tag{5.8}$$

(ii) The upper bound of $\psi_1^{(\mathrm{DT})}(x, \xi^\star)$ *in* (5.7) *is achieved at the points of the optimum design.*

(iii) For any non-optimum design ξ*, that is, a design for which* $\Phi_1^{(\mathrm{DT})}(\xi) < \Phi_1^{(\mathrm{DT})}(\xi^\star_{\mathrm{DT}})$,

$$\sup_{x \in X} \psi_1^{(\mathrm{DT})}(\xi) > 1. \tag{5.9}$$

Note that Theorem 5.1 is valid on the assumption that the answering set

$$\widehat{\Theta}_2(\xi^\star_{\mathrm{DT}}) = \mathrm{Arg} \min_{\vartheta_2 \in \Theta_2} \left\{ \int_X \{\eta(x) - \eta_2(x, \vartheta_2)\}^2 \, \xi^\star_{\mathrm{DT}}(\mathrm{d}x) \right\} \tag{5.10}$$

consists of a unique element, i.e., it is a singleton. If this is not the case, then a necessary and sufficient condition for $\xi^\star_{\mathrm{DT}}$ to be DT-optimal is the existence of a probability measure ζ defined on the sigma-algebra of Borel subsets of $\widehat{\Theta}_2(\xi^\star_{\mathrm{DT}})$ such that

$$\psi_1^{(\mathrm{DT})}(x, \xi^\star_{\mathrm{DT}}) \leq 1, \quad \forall x \in X, \tag{5.11}$$

where

$$\psi_1^{(\mathrm{DT})}(x, \xi) = \frac{1 - \kappa}{\Delta_1(\xi)} \int_{\widehat{\Theta}_2(\xi)} \left\{\eta(x) - \eta_2(x, \vartheta_2)\right\}^2 \zeta(\mathrm{d}\vartheta_2) + \frac{\kappa}{m_1} f_1^{\mathrm{T}}(x) M_1^{-1}(\xi) f_1(x). \tag{5.12}$$

In what follows, we wish to adopt the Wynn-Fedorov algorithm to numerically determine DT-optimal designs. For convenience, let us introduce the following notation: $\mathbb{R}_+$ stands for the set of nonnegative real numbers, and $\mathbb{R}_{++}$ stands for the set of positive real numbers. We will use the symbol $\mathrm{Sym}(n)$ to denote the set of symmetric $n \times n$ matrices, $\mathrm{NND}(n)$ to denote the set of symmetric nonnegative definite $n \times n$ matrices, and $\mathrm{PD}(n)$ to denote the set of symmetric positive-definite $n \times n$ matrices. Given $A, B \in \mathrm{Sym}(n)$, the notation $A \succeq B$ (resp. $A \succ B$) means that $A - B \in \mathrm{NND}(n)$ (resp. $A - B \in \mathrm{PD}(n)$).

5.1.1. Numerical construction of DT-optimum designs

Theorem 5.1 provides us with a test for optimality and it is employed to check whether or not an intuitively sensible design measure is DT-optimal. In practice, we require more than this, i.e., we need algorithms which enable us to construct DT-optimal design measures. In what follows, we limit ourselves to an adaptation of the Wynn-Fedorov scheme whose structure is as follows:

1 Guess an initial design ξ^0 such that $\Phi_1^{(\mathrm{DT})}(\xi^0) > -\infty$.

2 Iteratively compute a sequence of designs $\{\xi^k\}$ weakly converging to $\xi^\star$ as $k \to \infty$, the design ξ^{k+1} being obtained by a small perturbation of the design ξ^k, while still requiring $\Phi_1^{(\mathrm{DT})}(\xi^{k+1}) > -\infty$.

3 Stop computing $\{\xi^k\}$ after a finite number of steps if ξ^k can be qualified as almost optimal, e.g., in accordance with Theorem 5.1.

A grave difficulty encountered while trying to adapt those results in our case lies in the maximin form of the DT-optimality criterion.

The aim of this section is to study the convergence of the following procedure:

Algorithm 5.1 (Iterative construction of DT-optimum designs)**.**

Step 1: Guess a discrete starting design measure ξ^0 satisfying $\Delta_1(\xi^0) > 0$ and $\det(M_1(\xi^0)) > 0$. Choose some positive tolerance $\epsilon \ll 1$. Set $k = 0$.

Step 2: Find

$$\widehat{\vartheta}_2^k = \arg \min_{\vartheta_2 \in \Theta_2} \int_X \{\eta(x) - \eta_2(x, \vartheta_2)\}^2 \, \xi^k(\mathrm{d}x). \tag{5.13}$$

Step 3: Determine

$$\widehat{x}^k = \arg \max_{x \in X} \psi_1^{(\mathrm{DT})}(x, \xi^k), \tag{5.14}$$

where

$$\psi_1^{(\mathrm{DT})}(x, \xi^k) = \frac{1-\kappa}{\Delta_1(\xi^k)} \{\eta(x) - \eta_2(x, \widehat{\vartheta}_2^k)\}^2 + \frac{\kappa}{m_1} f_1^{\mathrm{T}}(x) M_1^{-1}(\xi^k) f_1(x). \tag{5.15}$$

If $\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) \leq 1 + \epsilon$, then STOP.

Step 4: For an appropriate value of $0 < \alpha^k < 1$, set

$$\xi^{k+1} = (1 - \alpha^k)\xi^k + \alpha^k \delta(\widehat{x}^k), \tag{5.16}$$

where $\delta(x)$ stands for the unit-weight design concentrated at x. Increment k by one and go to Step 2.

We can state these steps simply as follows:

1 Select any nondegenerate starting design.

2 Compute the parameter which minimizes the squared difference between the true and predicted responses integrated with respect to the current design.

3 Find a point of maximum sensitivity $\psi_1^{(\mathrm{DT})}(\,\cdot\,, \xi^k)$

4 Add the point of maximum sensitivity to the design with measure proportional to its sensitivity.

Note that for $\epsilon = 0$ the criterion for terminating the iterations, cf. Step 3, is nothing but the optimality condition of Theorem 5.1. Since, theoretically, in this case the above method is not finitely convergent (i.e., it is impossible to have $\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) \leq 1$ after a finite number of iterations), setting ϵ as a small positive scalar assures that we terminate reasonably close to an optimal design.

Step 4 raises the question of how to appropriately select the step size α^k. In the sequel, we shall consider the following choices:

(i) Limited maximization rule (Fedorov's algorithm):

$$\alpha^k = \arg \max_{\alpha \in [0,1]} \Phi_1^{(\mathrm{DT})}((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k)), \tag{5.17}$$

(ii) Armijo rule: Given fixed scalars $\beta, \sigma, s \in (0,1)$, we set

$$\alpha^k = \beta^{r_k} s, \tag{5.18}$$

where r_k is the first nonnegative integer r for which

$$\Phi_1^{(\mathrm{DT})}((1-\beta^r s)\xi^k + \beta^r s\delta(\widehat{x}^k)) - \Phi_1^{(\mathrm{DT})}(\xi^k) \geq \sigma\beta^r s\Big(\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\Big). \tag{5.19}$$

Usually σ is chosen close to zero (e.g., $\sigma \in [10^{-5}, 10^{-1}]$) and the reduction factor β is usually chosen from 1/2 to 1/10, cf. (Bertsekas, 1999, p. 29) for details. The directional differentiability of $\Phi_1^{(\mathrm{DT})}(\,\cdot\,)$ guarantees that this stepsize is well defined and will be found after a finite number of trial evaluations (5.19).

Further, we present our main results regarding convergence, for which we assume the following:

(B1) X and Θ_2 are compact sets.

(B2) $\eta(\,\cdot\,)$ and $f_1(\,\cdot\,)$ are continuous functions on X.

(B3) $\eta_2(\,\cdot\,,\,\cdot\,)$ is a continuous function on $X \times \Theta_2$.

Here and subsequently, we will use the symbol $\Xi(X)$ to denote the set of all probability measures on $\mathcal{B}(X)$, the σ-algebra of Borel subsets of X. To shorten notation, we let $e(x, \vartheta_2)$ stand for $\{\eta(x) - \eta_2(x, \vartheta_2)\}^2$ and write $\widehat{\Theta}_2(\xi)$ instead of $\operatorname{Arg}\min_{\vartheta_2 \in \Theta_2} \int_X e(x, \vartheta_2)\,\xi(\mathrm{d}x)$. It is easily seen that e is continuous and $\widehat{\Theta}_2(\xi)$ is compact. Similarly, we abbreviate $\arg\min_{\vartheta_2 \in \Theta_2} \int_X e(x, \vartheta_2)\,\xi(\mathrm{d}x)$ to $\widehat{\vartheta}_2(\xi)$.

Our convergence result for Algorithm 5.1 relies on the following lemma:

Lemma 5.1. *Let $\{\xi^k\}$ be an arbitrary sequence of designs from $\Xi(X)$ such that $\Phi_1^{(\mathrm{DT})}(\xi^k) \to c \in \mathbb{R}$ as $k \to \infty$. Then there is a design $\bar{\xi} \in \Xi(X)$ satisfying $c = \Phi_1^{(\mathrm{DT})}(\bar{\xi})$. Moreover, there exists a subsequence $\{\xi^k\}_{\mathcal{K}}$ of $\{\xi^k\}$, $\mathcal{K} \subset \mathbb{N}$, satisfying the following properties:*

(i) $\{\xi^k\}_{\mathcal{K}} \to \bar{\xi}$ *weakly.*

(ii) $\{\widehat{\vartheta}_2^k\}_{\mathcal{K}} \to \bar{\vartheta}_2$ *for some choice of minimizers* $\widehat{\vartheta}_2^k \in \widehat{\Theta}_2(\xi^k)$, $k \in \mathcal{K}$ *and* $\bar{\vartheta}_2 \in \widehat{\Theta}_2(\bar{\xi})$.

(iii) $\{M_1(\xi^k)\}_{\mathcal{K}} \to M_1(\bar{\xi}) \succ 0$.

(iv) $\{\Delta_1(\xi^k)\}_{\mathcal{K}} \to \Delta_1(\bar{\xi}) > 0$.

(v) $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}_{\mathcal{K}} \to \Phi_1^{(\mathrm{DT})}(\bar{\xi})$.

(vi) Furthermore, if $\widehat{\Theta}_2(\xi^k)$, $k \in \mathcal{K}$ *and* $\widehat{\Theta}_2(\bar{\xi})$ *are singletons, then*

$$\{\psi_1^{(\mathrm{DT})}(\widehat{x}(\xi^k), \xi^k)\}_{\mathcal{K}} \to \psi_1^{(\mathrm{DT})}(\widehat{x}(\bar{\xi}), \bar{\xi}) \begin{cases} = 1 & \text{if } \bar{\xi} \text{ is DT-optimal,} \\ > 1 & \text{otherwise,} \end{cases} \tag{5.20}$$

where

$$\widehat{x}(\xi) = \arg\max_{x \in X} \psi_1^{(\mathrm{DT})}(x, \xi). \tag{5.21}$$

Proof. First, observe that, possibly up to a finite number of terms, the sequence $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}$ is bounded because it is convergent. Therefore, we may assume that for each $k = 1, 2, \ldots$ we have $\Delta_1(\xi^k) > 0$ and $M_1(\xi^k) \succ 0$.

Consider the sequence $\{\widehat{\vartheta}_2^k\}$, where $\widehat{\vartheta}_2^k$ is any fixed element of $\widehat{\Theta}_2(\xi^k)$. Since Θ_2 is compact, it follows that there is a subsequence $\{\widehat{\vartheta}_2^k\}_{\mathcal{K}^0}$ of $\{\widehat{\vartheta}_2^k\}$, $\mathcal{K}^0 \subset \mathbb{N}$, which converges to some $\bar{\vartheta}_2 \in \Theta_2$. Furthermore, from Lemma 3.1, we deduce that there must be a subsequence $\{\xi^k\}_{\mathcal{K}}$ of $\{\xi^k\}_{\mathcal{K}^0}$, $\mathcal{K} \subset \mathcal{K}^0$, which convergences weakly to some $\bar{\xi} \in \Xi(X)$.

Consequently, Definition 3.1 clearly forces

$$\left\{\int_X e(x, \bar{\vartheta}_2)\, \xi^k(\mathrm{d}x)\right\}_{\mathcal{K}} \to \int_X e(x, \bar{\vartheta}_2)\, \xi(\mathrm{d}x) \tag{5.22}$$

and

$$\{M_1(\xi^k)\}_{\mathcal{K}} \to M_1(\bar{\xi}). \tag{5.23}$$

By the weak continuity of $\Delta_1(\,\cdot\,)$, cf. Theorem 3.1, we must also have

$$\{\Delta_1(\xi^k)\}_{\mathcal{K}} \to \Delta_1(\bar{\xi}). \tag{5.24}$$

On account of (5.23) and the continuity of the determinant, we thus have

$$\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}_{\mathcal{K}} \to \Phi_1^{(\mathrm{DT})}(\bar{\xi}). \tag{5.25}$$

Obviously, neither $\Delta_1(\bar{\xi})$ nor $\det\big(M_1(\bar{\xi})\big)$ is zero, since otherwise we would have $\Phi_1^{(\mathrm{DT})}(\bar{\xi}) = -\infty$, which is impossible.

It remains to show Part (vi). To this end, assume that $\widehat{\Theta}_2(\xi^k)$, $k \in \mathcal{K}$ and $\widehat{\Theta}_2(\bar{\xi})$ are singletons. Define

$$q(\vartheta_2, r, D) = \max_{x \in X} v(x, \vartheta_2, r, D), \tag{5.26}$$

where

$$v(x, \vartheta_2, r, D) = \frac{1 - \kappa}{r} e(x, \vartheta_2) + \frac{\kappa}{p_1} f_1^{\mathrm{T}}(x) D f_1(x). \tag{5.27}$$

Since v is continuous on $X \times \Theta_2 \times \mathbb{R}_{++} \times \mathrm{PD}(p_1)$, Corollary 5.4.2 of (Polak, 1997, p. 683) shows that q is continuous on $\Theta_2 \times \mathbb{R}_{++} \times \mathrm{PD}(p_1)$. Thus

$$\psi_1^{(\mathrm{DT})}(\widehat{x}(\xi^k), \xi^k) = q(\widehat{\vartheta}_2^k, \Delta_1(\xi^k), M_1^{-1}(\xi^k)), \qquad \forall k \in \mathcal{K} \tag{5.28}$$

and, since

$$\big\{\widehat{\vartheta}_2^k\big\}_{\mathcal{K}} \to \bar{\vartheta}_2, \qquad \big\{\Delta_1(\xi^k)\big\}_{\mathcal{K}} \to \Delta_1(\bar{\xi}), \qquad \big\{M_1^{-1}(\xi^k)\big\}_{\mathcal{K}} \to M_1^{-1}(\bar{\xi}), \tag{5.29}$$

we get

$$\big\{\psi_1^{(\mathrm{DT})}(\widehat{x}(\xi^k), \xi^k)\big\}_{\mathcal{K}} \to \psi_1^{(\mathrm{DT})}(\widehat{x}(\bar{\xi}), \bar{\xi}). \tag{5.30}$$

The method of choosing $\widehat{x}(\bar{\xi})$ implies that

$$\begin{aligned} \psi_1^{(\mathrm{DT})}(\widehat{x}(\bar{\xi}), \bar{\xi}) - 1 &= \left.\frac{\partial \Phi_1^{(\mathrm{DT})}\big((1-\alpha)\bar{\xi} + \alpha\delta(\widehat{x}(\bar{\xi}))\big)}{\partial \alpha}\right|_{\alpha=0^+} \\ &= \lim_{\alpha \downarrow 0} \frac{1}{\alpha}\Big(\Phi_1^{(\mathrm{DT})}\big((1-\alpha)\bar{\xi} + \alpha\delta(\widehat{x}(\bar{\xi}))\big) - \Phi_1^{(\mathrm{DT})}(\bar{\xi})\Big) \geq 0. \end{aligned} \tag{5.31}$$

If $\psi_1^{(\mathrm{DT})}(\widehat{x}(\bar{\xi}), \bar{\xi}) = 1$, then $\bar{\xi}$ must be optimal, cf. Theorem 5.1. For a nonoptimal $\bar{\xi}$ we thus get $\psi_1^{(\mathrm{DT})}(\widehat{x}(\bar{\xi}), \bar{\xi}) > 1$. This completes the proof. □

We can now formulate our main result regarding convergence.

Theorem 5.2. *Let $\{\xi^k\}$ be a sequence of designs generated by Algorithm 5.1 with α^k chosen by the Armijo rule. Morever, assume that the sets $\widehat{\Theta}_2((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k))$ are singletons for each $k = 0, 1, \ldots$ and for all $\alpha \in [0, s]$, $s < 1$. Then the sequence $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}$ is nondecreasing and converges to $\max\{\Phi_1^{(\mathrm{DT})}(\xi) : \xi \in \Xi(X)\}$.*

Proof. Based on Lemma 3.1, it is easy to check that the set $\{M_1(\xi) : \xi \in \Xi(X)\}$ is compact. This, taken in conjunction with the continuity of $e(\,\cdot\,,\,\cdot\,)$ on $X \times \Theta_2$, implies that $\Phi_1^{(\mathrm{DT})}(\xi) < \infty$ for all $\xi \in \Xi(X)$. Consequently, $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}$ is bounded.

With the notation $\widehat{x}^k = \arg\max_{x \in X} \psi_1^{(\mathrm{DT})}(x, \xi^k)$, the definition of the Armijo rule yields

$$\Phi_1^{(\mathrm{DT})}(\xi^{k+1}) - \Phi_1^{(\mathrm{DT})}(\xi^k) \geq \sigma\alpha^k\Big(\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\Big), \qquad k = 0, 1, \ldots \tag{5.32}$$

and hence the sequence $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}$ is monotonically nondecreasing and thus it converges to a finite value c.

It is clear that for each $k = 0, 1, \ldots$ there must be

$$\left.\frac{\partial \Phi_1^{(\mathrm{DT})}((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k))}{\partial \alpha}\right|_{\alpha=0^+} = \psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1 \geq 0. \tag{5.33}$$

Since $\psi_1^{(\mathrm{DT})}(\widehat{x}^{k_0}, \xi^{k_0}) = 1$ for some index k_0 means that ξ^{k_0} is a DT-optimum design and the computation can be interrupted, we can assume that for each $k = 0, 1, \ldots$ we have $\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) > 1$.

For the sake of contradiction, assume that $c < \max\{\Phi_1^{(\mathrm{DT})}(\xi) : \xi \in \Xi(X)\}$. From Lemma 5.1 we know that $c = \Phi_1^{(\mathrm{DT})}(\bar{\xi})$ for some $\bar{\xi} \in \Xi(X)$ and there exists a subsequence $\{\xi^k\}_{\mathcal{K}}$, $\mathcal{K} \subset \mathbb{N}$, which converges weakly to $\bar{\xi}$.

Since $\{\Phi_1^{(\mathrm{DT})}(\xi^k)\}$ is convergent, it follows that

$$\left\{\Phi_1^{(\mathrm{DT})}(\xi^{k+1}) - \Phi_1^{(\mathrm{DT})}(\xi^k)\right\} \to 0. \tag{5.34}$$

Conseqently, by the Armijo rule,

$$\left\{\alpha^k\left(\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\right)\right\} \to 0. \tag{5.35}$$

Moreover,

$$\left\{\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\right\}_{\mathcal{K}} \to \omega > 0 \tag{5.36}$$

because $\bar{\xi}$ is nonoptimal, cf. Part (vi) of Lemma 5.1.

Hence

$$\{\alpha^k\}_{\mathcal{K}} \to 0. \tag{5.37}$$

Therefore, by the definition of the Armijo rule, for some index $k' \geq 0$ we must have

$$\Phi_1^{(\mathrm{DT})}((1-\bar{\alpha}^k)\xi^k + \bar{\alpha}^k\delta(\widehat{x}^k)) - \Phi_1^{(\mathrm{DT})}(\xi^k) < \sigma\bar{\alpha}^k\left(\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\right), \quad \forall k \in \mathcal{K}, \quad k \geq k', \tag{5.38}$$

where $\bar{\alpha}^k = \alpha^k/\beta$. In practice, this means that the initial stepsize s will be reduced at least once for all $k \in \mathcal{K}$, $k \geq k'$. If there were no such reductions, this would contradict (5.37).

We get

$$\frac{1}{\bar{\alpha}^k}\left(\Phi_1^{(\mathrm{DT})}((1-\bar{\alpha}^k)\xi^k + \bar{\alpha}^k\delta(\widehat{x}^k)) - \Phi_1^{(\mathrm{DT})}(\xi^k)\right) < \sigma\left(\psi_1^{(\mathrm{DT})}(\widehat{x}^k, \xi^k) - 1\right), \qquad \forall k \in \mathcal{K}, \quad k \geq k'. \tag{5.39}$$

Define

$$g^k(\alpha) = (1-\kappa)\ln\left(\Delta_1((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k))\right) + \frac{\kappa}{p_1}\ln\left(\det(M_1((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k)))\right). \tag{5.40}$$

We claim that under the assumptions of the theorem, g is continously differentiable. Indeed, there is no problem with the second term on the right-hand side of (5.40) and the only tricky question is the differentiability of

$$g_1^k(\alpha) = \Delta_1((1-\alpha)\xi^k + \alpha\delta(\widehat{x}^k)) = \min_{\vartheta_2 \in \Theta_2} \varrho^k(\alpha, \vartheta_2), \tag{5.41}$$

where

$$\varrho^k(\alpha,\vartheta_2) = (1-\alpha)\int_X e(x,\vartheta_2)\,\xi^k(\mathrm{d}x) + \alpha e(\widehat{x}^k,\vartheta_2). \tag{5.42}$$

Writing

$$\widetilde{\Theta}_2^k(\alpha) = \operatorname{Arg}\min_{\vartheta_2\in\Theta_2} \varrho^k(\alpha,\vartheta_2), \tag{5.43}$$

we conclude from Danskin's theorem (Bertsekas, 1999, Prop. B.25, p. 717) that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} g_1^k(\alpha) = \min_{\vartheta_2\in\widetilde{\Theta}_2^k(\alpha)} \frac{\partial\varrho^k(\alpha,\vartheta_2)}{\partial\alpha}. \tag{5.44}$$

But on account of the valid assumptions, $\widetilde{\Theta}_2^k(\alpha)$ is a singleton, i.e.,

$$\widetilde{\Theta}_2^k(\alpha) = \{\widetilde{\vartheta}_2^k(\alpha)\}, \tag{5.45}$$

which gives

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} g_1^k(\alpha) = \frac{\partial\varrho^k(\alpha,\widetilde{\vartheta}_2^k(\alpha))}{\partial\alpha} = e(\widehat{x}^k,\widetilde{\vartheta}_2^k(\alpha)) - \int_X e(x,\widetilde{\vartheta}_2^k(\alpha))\,\xi^k(\mathrm{d}x). \tag{5.46}$$

Since $\widetilde{\vartheta}_2^k(\,\cdot\,)$ is continuous on $[0,1]$, cf. Theorem 5.4.3 of (Polak, 1997, p. 684), it follows that g_1^k is continuously differentiable, and consequently, so is g^k. Therefore we can apply the Mean Value Theorem to get

$$\frac{g^k(\bar{\alpha}^k) - g^k(0)}{\bar{\alpha}^k} = \frac{\mathrm{d}g^k}{\mathrm{d}\alpha}(\breve{\alpha}^k) \tag{5.47}$$

for some $\breve{\alpha}^k \in (0,\bar{\alpha}^k)$, i.e.,

$$\begin{aligned}
&\frac{1}{\bar{\alpha}^k}\Big(\Phi_1^{(\mathrm{DT})}((1-\bar{\alpha}^k)\xi^k + \bar{\alpha}^k\delta(\widehat{x}^k)) - \Phi_1^{(\mathrm{DT})}(\xi^k)\Big)\\
&\quad= \frac{(1-\kappa)}{\Delta_1((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))}\left(e(\widehat{x}^k,\widetilde{\vartheta}_2^k(\breve{\alpha}^k)) - \int_X e(x,\widetilde{\vartheta}_2^k(\breve{\alpha}^k))\,\xi^k(\mathrm{d}x)\right)\\
&\qquad+ \frac{\kappa}{m_1}\Big(f_1^{\mathrm{T}}(\widehat{x}^k)M_1^{-1}((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))f_1(\widehat{x}^k)\\
&\qquad\qquad- \operatorname{trace}\big\{M_1^{-1}((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))M_1(\xi^k)\big\}\Big)
\end{aligned} \tag{5.48}$$

From (5.39) we see that

$$\begin{aligned}
&\frac{(1-\kappa)}{\Delta_1((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))}\left(e(\widehat{x}^k,\widetilde{\vartheta}_2^k(\breve{\alpha}^k)) - \int_X e(x,\widetilde{\vartheta}_2^k(\breve{\alpha}^k))\,\xi^k(\mathrm{d}x)\right)\\
&\qquad+ \frac{\kappa}{p_1}\Big(f_1^{\mathrm{T}}(\widehat{x}^k)M_1^{-1}((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))f_1(\widehat{x}^k)\\
&\qquad\qquad- \operatorname{trace}\big\{M_1^{-1}((1-\breve{\alpha}^k)\xi^k + \breve{\alpha}^k\delta(\widehat{x}^k))M_1(\xi^k)\big\}\Big)\\
&\qquad\qquad< \sigma\Big(\psi_1^{(\mathrm{DT})}(\widehat{x}^k,\xi^k) - 1\Big), \qquad \forall k\in\mathcal{K},\quad k\geq k'.
\end{aligned} \tag{5.49}$$

It is an easy exercise to check that, as $k \to^{\mathcal{K}} \infty$, we finally get

$$\omega \leq \sigma\omega \tag{5.50}$$

or

$$(1-\sigma)\omega \leq 0. \tag{5.51}$$

But $\sigma < 1$, and hence

$$\omega \leq 0. \tag{5.52}$$

This contradicts (5.36), and the theorem is proved. □

The proof of the convergence for the limited maximization rule proceeds on the same lines. The following numerical example serves as a vehicle for the display of some salient features of the above solution technique.

Example 5.1. Let consider the process of chemical conversion of substance A into B and C in a batch reactor. Two rival models are considered for this system:

Model 1: A reversible first-order reaction followed by an irreversible reaction $A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \overset{k_2}{\to} C$, which is described by the equations

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A] + k_3[B], \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A] - k_2[B]^{\lambda} - k_3[B], \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B]^{\lambda}, \\ [A]_{t=0} &= a_0, \quad [B]_{t=0} = b_0, \quad [C]_{t=0} = c_0, \end{aligned} \tag{5.53}$$

where $[A]$, $[B]$ and $[C]$ denote concentrations of reagents A, B and C, respectively, and a_0, b_0, c_0 stand for initial concentrations.

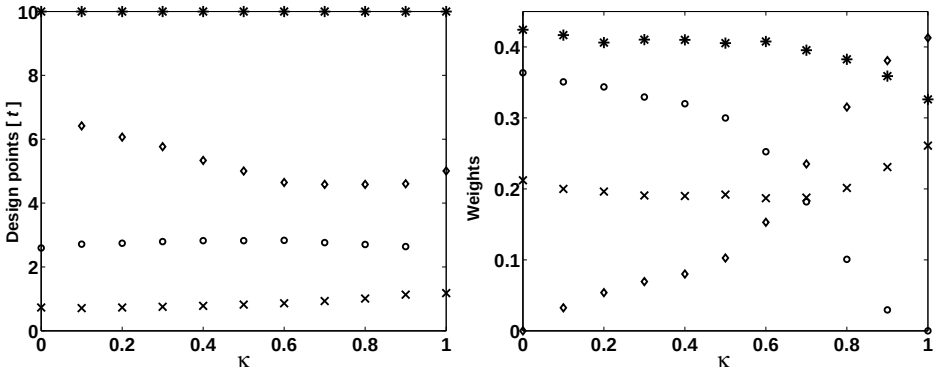
Model 2: Consecutive irreversible reactions $A \overset{k_1}{\to} B \overset{k_2}{\to} C$ described by the equations

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A]^{\lambda_1}, \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A]^{\lambda_1} - k_2[B]^{\lambda_2}, \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B]^{\lambda_2}, \\ [A]_{t=0} &= a_0, \quad [B]_{t=0} = b_0, \quad [C]_{t=0} = c_0. \end{aligned} \tag{5.54}$$

Model 1 is assumed to be the true one with parameters $(k_1, k_2, k_3, \lambda) = (0.7, 0.2, 0.1, 2.0)$. The acceptable range of parameters of Model 2 which is considered alternative is set as $0.55 \leq k_1 \leq 0.85$, $0.05 \leq k_2 \leq 0.35$, $1.5 \leq \lambda_1 \leq 2.5$,

$1.5 \leq \lambda_2 \leq 2.5$. The time horizon (the design region) and initial concentrations are respectively set to $T = [0, 10]$ and $x_0 = (a_0, b_0, c_0) = (1, 0, 0)$ for both the models.

In order to determine DT-optimum designs, a program was written using the Lahey-Fujitsu Fortran 95 compiler v.5.6 with IMSL 4.0 library. The attendant problems of global optimization were solved using the ARS scheme supplemented with a local optimizer (modified Levenberg-Marquardt algorithm) from the IMSL library (`DBCLSJ` subroutine). The obtained results are presented in Figs. 5.1 and 5.2. Figure 5.3 shows the plot of an exemplary sensitivity function obtained for $\kappa = 0.5$.



(a) (b)

Fig. 5.1. Structure of designs as κ varies: (a) support points t_i, (b) corresponding design weights w_i. There are four points of support, except of three-point designs for $\kappa = 0$ and $\kappa = 1$, the T-optimum and D-optimum designs, respectively.

5.2. T-optimum designs on finite design spaces

The basic assumption in this section is that the set of admissible support points X, where the observations are possible, is finite, i.e., $X = \{x_1, \ldots, x_n\}$ where $x_i,\ i = 1, \ldots, n$ are given *a priori*. Because the allowable measurement points are fixed, any design $\xi \in \Xi(X)$ has the form

$$\xi = \begin{Bmatrix} x_1, & \ldots, & x_n \\ p_1, & \ldots, & p_n \end{Bmatrix} \tag{5.55}$$

and is uniquely determined by the values of the weights $p_1, \ldots, p_n$. Thus, the problem simply reduces to the optimization of the design weights and it will cause no confusion if we use the symbol $\Delta_1(p)$ to designate the value of the genuine T-optimality criterion $\Delta_1(\xi)$ at the design of the form (5.55).

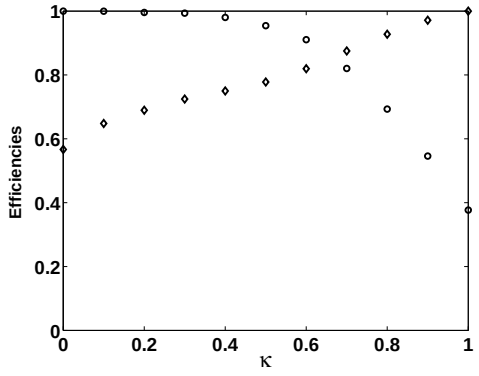


Fig. 5.2. Efficiencies of designs as κ varies: $\circ$ T-efficiency; $\diamond$ D-efficiency. The D-efficiency increases as κ increases, whereas the T-efficiency decreases.

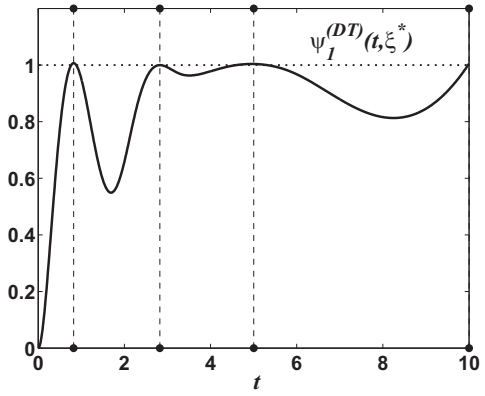


Fig. 5.3. Exemplary sensitivity function in Example 5.1 for $\kappa = 0.5$. Vertical dashed lines reflect the location of the optimal support points (the fourth point is just the end of the observation horizon).

In this nomenclature, the problem may be stated as follows: Find the weight vector $p \in \mathbb{R}^n$ that maximizes

$$\Delta_1(p) = \min_{\vartheta_2 \in \Theta_2} \mathcal{J}(p, \vartheta_2), \tag{5.56}$$

where

$$\mathcal{J}(p, \vartheta_2) = \sum_{i=1}^{n} p_i \|\eta(x_i) - \eta_2(x_i, \vartheta_2)\|^2, \tag{5.57}$$

subject to

$$p \in \mathbb{S} = \Big\{ p = (p_1, \ldots, p_n) : \; p_i \geq 0, \; i = 1, \ldots, n; \quad \sum_{i=1}^{n} p_i = 1 \Big\}. \tag{5.58}$$

This is a finite-dimensional optimization problem over the canonical simplex $\mathbb{S}$.

5.2.1. Gradient projection algorithm

Obviously, it is always possible to exploit some general constrained optimization approaches in order to solve the problem formulated above (cf. the next section). However, due to a relatively simple form of the constraints, a more straightforward procedure can be proposed, which reduces to using a gradient projection method (Bertsekas, 1999, p. 223). The first step is to find a feasible direction, i.e., the one which guarantees an increase in the value of the criterion $\Delta_1(\,\cdot\,)$ and then a step is taken along this line. The result is projected on $\mathbb{S}$, thereby obtaining a new feasible weight vector.

Generation of a new candidate point can be formalized as follows:

$$p_+^{k+1} = \Pi_{\mathbb{S}}\Big[p^k + \alpha^k d^k \Big], \tag{5.59}$$

where $d^k \in \mathbb{R}^n$ is the vector representing a feasible direction of weight modification, α^k is some positive coefficient which controls the correction process, and $\Pi_{\mathbb{S}}[\,\cdot\,]$ stands for the orthogonal projection onto the canonical simplex $\mathbb{S}$.

If the answering set $\widehat{\Theta}_2(p^k) = \operatorname{Arg} \min\limits_{\vartheta_2 \in \Theta_2} \mathcal{J}(p, \vartheta_2) = \big\{ \vartheta_2^k \big\}$, i.e., it is a singleton, then by Danskin's Theorem (Bertsekas, 1999, Prop. B.25, p. 717) the criterion $\Delta_1(\,\cdot\,)$ is differentiable at p^k and the derivatives

$$\frac{\partial \Delta_1(p^k)}{\partial p_i} = \|\eta(x_i) - \eta_2(x_i, \vartheta_2^k)\|^2, \quad i = 1, \ldots, n \tag{5.60}$$

are extremely easy to calculate, so the very first idea is to choose the direction d^k determined by the gradient $\nabla\big(\Delta_1(p)\big)\big|_{p=p^k}$ as in steepest ascent. Also note that there exist many possible choices of determining the step coefficient α^k. It can be taken as a suitable constant. Also its value can be adjusted according to

$$\alpha_k = \arg \max_{\alpha > 0} \Delta_1\big(\Pi_{\mathbb{S}}\big(p^k + \alpha d^k\big)\big) \tag{5.61}$$

However, the computational effort to determine such an optimum value can easily outgrow potential benefits.

A reasonable approach consist in an adaptive line search using the Armijo rule (Bertsekas, 1999). Here, fixed scalars s, β and σ, with $0 < \beta < 1$, and $0 < \sigma < 1$ are chosen, and we set $\alpha^k = \beta^{r_k} s$, where r_k is the first nonnegative integer r for which

$$\Delta_1\big(p^k + \beta^r s d^k\big) - \Delta_1(p^k) \geq \sigma \beta^r s \frac{\partial \Delta_1(p^k)}{\partial p} d^k. \tag{5.62}$$

Finally, we should remind that each iteration of the rule (5.59) must be preceded with calculation of ϑ_2^k, which implies a global minimization process.

Then, summarizing, the following steepest-ascent type algorithm can be developed:

Algorithm 5.2 (Weight optimization algorithm for T-optimum designs, v.1)**.**

Step 1: Guess a starting set of weights $p^0 \in \mathbb{S}$. Choose some positive tolerance $\epsilon \ll 1$. Set $k = 0$.

Step 2: Compute

$$\widehat{\vartheta}_2^k = \arg \min_{\vartheta_2 \in \Theta_2} \sum_{i=1}^{n} p_i^k \|\eta(x_i) - \eta_2(x_i, \vartheta_2)\|^2 \tag{5.63}$$

and

$$p^{k+1} = \Pi_{\mathbb{S}}\Big[p^k + \alpha_k d^k\Big] \tag{5.64}$$

where

$$d^k = \begin{bmatrix} \|\eta(x_1) - \eta_2(x_1, \widehat{\vartheta}_2^k)\|^2 \\ \vdots \\ \|\eta(x_n) - \eta_2(x_n, \widehat{\vartheta}_2^k)\|^2 \end{bmatrix} \tag{5.65}$$

and α^k is computed according to the rule (5.62).

Step 3: If the condition

$$\|p^{k+1} - p^k\| < \epsilon \tag{5.66}$$

is satisfied, then *STOP*. Otherwise, increment k by one and go to Step 2.

At first sight, the gradient projection procedure above is rather easy in implementation due to its simplicity, but unfortunately, it inherits all the drawbacks of steepest-ascent-like algorithms. For example, the convergence rate dramatically decreases in the vicinity of the minimum. This can be avoided to some extent with a suitable choice of the step length α^k (like in the proposed Armijo rule). As for projection onto the canonical simplex $\mathbb{S}$, an algorithm can be developed which is almost as simple as a closed-form solution. Indeed, the point $\widetilde{p} = \Pi_{\mathbb{S}}[p]$ is defined as the solution to the problem:
Minimize

$$\sum_{i=1}^{n} (p_i - \widetilde{p}_i)^2 \tag{5.67}$$

subject to

$$\widetilde{p}_i \geq 0, \quad i = 1, \ldots, n, \quad \sum_{i=1}^{n} \widetilde{p}_i = 1. \tag{5.68}$$

An elegant and simple algorithm for solving this task is described by (Tuenter, 2001). Without loss of generality, assume that $p_1 \geq p_2 \geq \cdots \geq p_n$, since this is only a matter of reordering the elements of p.

Finding a sought projection proceeds as follows:

Step 1: Set $q = 0$ and $\ell = 1$.

Step 2: Set $\ell^\star = \ell$ and $q^\star = q$. Increment ℓ by one and update $q \leftarrow q + \ell^\star(p_{\ell^\star} - p_\ell)$.

Step 3: If $q > 1$, then go to Step 5.

Step 4: If $\ell = n$ then set $\ell^\star = \ell$, $q^\star = q$ and go to Step 5. Otherwise, go to Step 2.

Step 5: Set

$$\lambda = \frac{1 - q^\star}{\ell^\star} - p_{\ell^\star}, \tag{5.69}$$

and the components of the sought projection $\widetilde{p}$ as follows:

$$\widetilde{p}_i = \begin{cases} p_i + \lambda & \text{for } 1 \leq i \leq \ell^\star, \\ 0 & \text{otherwise,} \end{cases} \tag{5.70}$$

STOP.

In spite of the simplicity of this algorithm, it may still involve significant numerical efforts. Moreover, the projection operator is not differentiable, and therefore the search for an optimum step-length α^k requires a derivative-free algorithm (e.g., the golden-search one). Moreover, the parameters in (5.62) have to be suitably chosen. Consequently, the effective usage of the proposed weight optimization algorithm requires practical experience from the user to overcome some impediments.

5.2.2. Adaptation of the Uzawa method

Given $\vartheta_2 \in \Theta_2$, the function $\mathcal{J}(\,\cdot\,, \vartheta_2)$ is linear and hence concave. Therefore the function $\Delta_1(\,\cdot\,)$ is concave since it is defined as the pointwise supremum over the infinite family of concave functions $\{\mathcal{J}(\,\cdot\,, \vartheta_2)\}_{\vartheta_2 \in \Theta_2}$ (Boyd and Vandenberghe, 2004). Its maximization over the canonical simplex can be formulated as follows: Find a weight vector $p \in \mathbb{R}^n$ that produces the maximum of

$$\Delta_1(p) = \sum_{i=1}^{n} p_i \|\eta(x_i) - \eta_2(x_i, \vartheta_2(p))\|^2$$

subject to

$$\begin{aligned} g_1(p) &= p_1 \geq 0, \\ &\vdots \\ g_{n-1}(p) &= p_{n-1} \geq 0, \\ g_n(p) &= 1 - \sum_{j=1}^{n-1} p_j \geq 0. \end{aligned} \tag{5.71}$$

To solve the problem (5.71) we propose an adaptation of the Uzawa method for nonlinear saddle-point problems (Chen, 1998). The resulting algorithm is as follows:

Algorithm 5.3 (Weight optimization algorithm for T-optimum designs, v.2)**.**

Construct the Lagrange function $L(p, \mu) = \Delta_1(p) + \sum_{j=1}^{n} \mu_j g_j(p)$.

Step 1: Guess starting values of Lagrange multipliers $\mu_j \geq 0$, $j = 1, \ldots, n$. Choose some positive tolerance $\epsilon \ll 1$ and a positive steplength ρ. Set $k = 0$.

Step 2: Find $p^k = \arg\max_{p \in \mathbb{R}^n} L(p, \mu^k)$.

Step 3: If $\sum_{j=1}^{n} \mu_j |g_j(p^k)| \leq \epsilon$ and $\min\left\{g_1(p^k), \ldots, g_n(p^k)\right\} \geq -\epsilon$ then STOP.

Step 4: Set $\mu_j^{k+1} = \max(0, \mu_j^k - \rho g_j(p^k))$, $k \leftarrow k + 1$ and go to Step 2.

Generally, the method consists in sequential relaxation with respect to weights and Lagrange multipliers. Stopping conditions result from the Karush-Kuhn-Tucker optimality conditions. Global convergence is guaranteed for a sufficiently small ρ.

Both the procedures outlined above can be incorporated into the Wynn-Fedorov scheme, which may significantly decrease the number of iterations necessitated to obtain an optimum design (since proper weight optimization takes much more time in W-F type procedures than the identification of optimum or sub-optimum support points). In order to check the usefulness of the presented approach, the discrimination problem from Example 4.2 was considered again with the same settings, but using Algorithm 5.2 to find a T-optimum design. For that purpose, $n = 100$ evenly spaced support points were chosen in the design domain $T = [0, 10]$. The initial weights p^0 were generated randomly. Parameters of the algorithm were $\epsilon = 0.0001$, $s = 25$, $\beta = 0.99$ and $\sigma = 0.2$. The code written in Fortran 95 was executed using the Lahey Fujitsu Fortran 95 compiler. The obtained solution (weights and the corresponding sensitivity function) is shown in Fig. 5.4. The algorithm converged in 140 iterations. Parameter values of the 'worst' alternative model are $\vartheta_2^\star = (0.8475, 0.1688, 1.5764, 1.7281)$. The results remain in full agreement with those of Example 4.2.

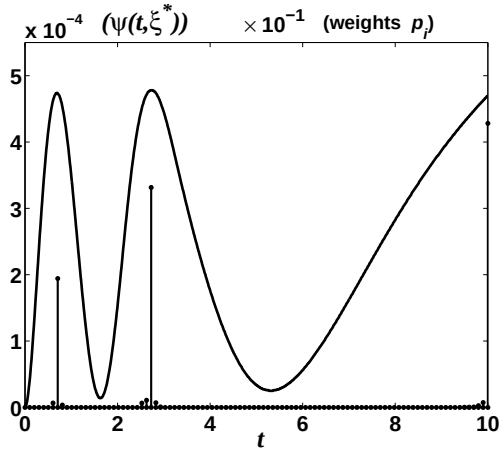


Fig. 5.4. Sensitivity function and weights of the obtained T-optimum design on finite support space (100 points uniformly distributed on T) for the discrimination problem of Example 4.2. Note that different scalings of the ordinates are used for the sensitivity and weights.

5.3. Selective random search for T-optimum experimental design

One of the main two phases of the Wynn-Fedorow procedure described in Section 4.1 consists in determination of a candidate point $x^k = \max_{x \in X} \psi(x, \xi^k)$, where $\psi(x, \xi^k) = \|\eta(x) - \eta_2(x, \widehat{\vartheta}_2^k)\|^2$. Instead of searching for an accurate maximum, which can be very costly in a high-dimensional case, we can limit ourselves to obtain a point which will be 'good enough' to increase the 'quality of the design', i.e., a point x_t for which $\psi(x_t, \xi^k) > \Delta_1(\xi^k)$ (see Fig. 5.5). Remark that from Theorem 3.8 it follows that such a point for a non-optimum design ξ^k exists. To develop an efficient method of generating of the candidate x_t, observe that $\psi(x, \xi^k)$ is a non-negative function. Furthermore,

$$f_k(x) = \frac{\psi(x, \xi^k)}{\int_X \psi(x, \xi^k)\,\mathrm{d}x} \tag{5.72}$$

defines the probability density function of a continuous random variable $\mathcal{X}$. This density has global maxima just at the points which are candidates for inclusion into the current design. Thus, if we can simulate the contrived random variable $\mathcal{X}$, their neighbours are presumably more likely to be observed in a generated sample than in a blind random search with uniform distribution over X or the ARS mechanism.

The above observation provides a basis for the so-called selective random search method set forth by Rafajłowicz to efficiently determine D-optimum designs (Rafajłowicz, 1998), where the convergence with probability one was also proved. Unfortunately, the computational effort due to the use of the rejection

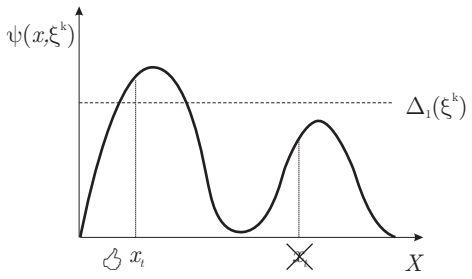


Fig. 5.5 . Idea of selective random search.

sampling suggested by Rafajłowicz may be still prohibitive since the complicated landscape of the function $\psi(\,\cdot\,,\xi^k)$, especially high multimodality, implies that the target density $f_k(\,\cdot\,)$ can be evaluated, but not easily sampled. To alleviate this inconvenience, we propose here to apply the Monte Carlo Markov Chain (MCMC) method (Givens and Hoeting, 2005; Gilks *et al.*, 1995). Generally, MCMC methods make use of the observation that it is possible to construct a suitable irreducible and aperiodic Markov chain (discrete or continuous) whose stationary distribution approximates some target density $f(\,\cdot\,)$. This means that such a chain can be used to generate a draw from a distribution that approximates $f(\,\cdot\,)$. The MCMC approach has several advantages, e.g., increasing problem dimensionality usually does not slow convergence or make the very simple implementation more complex. The main difficulty consists in construction of a chain with suitable properties. Since MCMC theory and applications are areas of active research, many algorithms and strategies have been proposed to achieve the most valuable behaviour of the resulting chain. Here we employ the classical and extremely simple Metropolis-Hastings algorithm. The method is as follows (Givens and Hoeting, 2005):

❶ Select $X^{(0)} = x^{(0)}$ drawn at random from some proposal distribution g such that $f_k\big(x^{(0)}\big) > 0$. Set $\ell = 0$.

❷ Given $X^{(\ell)} = x^{(\ell)}$ sample a candidate value X^c from a proposal distribution $g(\,\cdot\,|x^{(\ell)})$.

❸ Compute the Metropolis-Hastings ratio $R\big(x^{(\ell)}, X^c\big)$ where

$$R(a,b) = \frac{f_k(b)g(a|b)}{f_k(a)g(b|a)}\,. \tag{5.73}$$

❹ Sample a value

$$X^{(\ell+1)} = \begin{cases} X^c & \text{with probability } p = \min\{R(x^{(\ell)}, X^c), 1\}, \\ x^{(\ell)} & \text{otherwise.} \end{cases} \tag{5.74}$$

❺ Set $\ell \leftarrow \ell + 1$ and return to Step 2.

The crucial part of the method is the choice of a proposal distribution, since the properties of the obtained chain strongly depend on $g(\,\cdot\,|\,\cdot\,)$. It seems intuitively clear that we wish the proposal distribution to be easily sampled and to approximate the target distribution $f_k(\,\cdot\,)$ very well. However, if we do not possess information on such a good approximation, a reasonable choice is the uniform distribution over X. Then obviously $g(a|b) = g(b|a)$. For such a symmetric distribution the method is known as the Metropolis algorithm. Note that the unknown proportionality constant $\int_X \psi(x, \xi^k)\,\mathrm{d}x$ cancels in the numerator and denominator of (5.73), so that we then have

$$R(a, b) = \frac{\psi(b, \xi^k)}{\psi(a, \xi^k)}\,. \tag{5.75}$$

In what follows, we present our T-optimum equivalent to Rafajłowicz's selective random search:

Algorithm 5.4 (Selective random search for T-optimum designs)**.**

Step 1: Choose an initial nonsingular design ξ^0. Set $k = 0$.

Step 2: Find:

$$\widehat{\vartheta}_2^k = \arg \min_{\vartheta_2 \in \Theta_2} \sum_{i=1}^{n^k} p_i^k \|\eta(x_i^k) - \eta_2(x_i^k, \vartheta_2)\|^2, \tag{5.76}$$

where n^k is the number of support points in the current design, and then generate a candidate $\widehat{x}^k$ using the MCMC with the metropolis-Hastings ratio (5.75) and $g(\,\cdot\,)$ being the uniform distribution on X until

$$\psi(\widehat{x}^k, \xi^k) > \Delta_1(\xi^k), \tag{5.77}$$

where

$$\psi(x, \xi^k) = \|\eta(x) - \eta_2(x, \widehat{\vartheta}_2^k)\|^2, \tag{5.78}$$

$$\Delta(\xi^k) = \sum_{i=1}^{n^k} p_i^k \|\eta(x_i^k) - \eta_2(x_i^k, \widehat{\vartheta}_2^k)\|^2. \tag{5.79}$$

If the number of generated trial points exceeds some prescribed number $N_{\max}$ without satisfying (5.77) then STOP.

Step 3: Choose a suitable α^k with $0 \le \alpha^k \le 1$ and compute the convex combination of designs:

$$\xi^{k+1} = (1 - \alpha^k)\xi^k + \alpha^k \delta(\widehat{x}_k), \tag{5.80}$$

where $\delta(x)$ is the unit-weight design concentrated at x. Set $k \leftarrow k + 1$ and go to Step 2.

The only difference compared with the Wynn-Fedorov scheme appears in Step 2, when the above-mentioned concept of MCMC approximation to the maximizer of the sensitivity function is incorporated. The stopping condition results from the observation that if ξ^k is close to the optimum design or it is the optimum one, the condition (5.77) may never be satisfied. The value $N_{\max}$ should thus be suitably large in order to prevent the algorithm precociously terminating. The probability of such an event can be freely reduced by setting $N_{\max}$ adequately large, cf. (Rafajłowicz, 2006).

In order to verify the usefulness of the presented approach, a numerical approximation to the T-optimum design from Example 3.2 was sought using Algorithm 5.4. Since the analytical solution to this problem is known, we have an additional test for the quality of the found numerical solutions. Starting with a randomly chosen ten-point uniformly distributed (i.e., with weights equal to 1/10) initial design and $N_{\max} = 1000$, Algorithm 5.4 found the following approximation of the exact solution after 268 iterations (the values were rounded to the fourth fractional digit and some additional modifications of the basic procedure, like a removal of clusters and points with negligible weights, were also applied):

$$\xi^\star = \begin{Bmatrix} 0.0016, & 0.4984, & 0.9999 \\ 0.2555, & 0.4984, & 0.2461 \end{Bmatrix}. \tag{5.81}$$

The relevant computer program was implemented using Maple 10 as a set of Maple procedures. Figure 5.6 displays the resultant sensitivity function. The parameter values of the alternative model with respect to $\xi^\star$ rounded to the fifth fractional digit are $\vartheta_2^\star = (-0.12546, 0.99934)$.

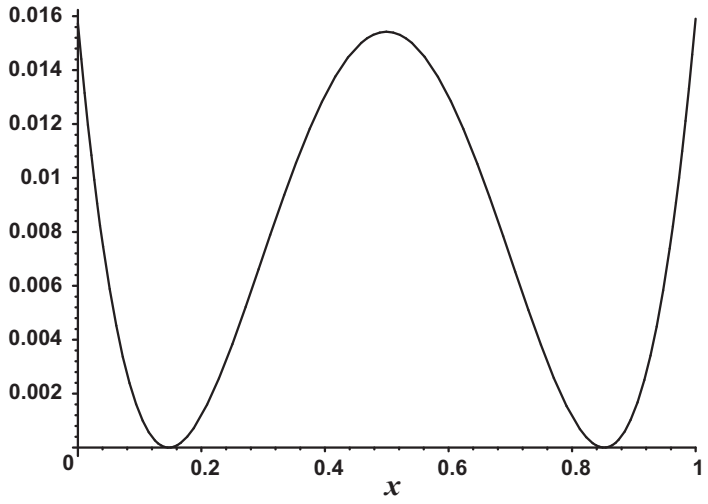


Fig. 5.6. Sensitivity function $\psi(x, \xi^\star)$ of Example 3.2 obtained using Algorithm 5.4.

5.4. Parallel computing for T-optimum designs

The algorithms delineated in the dissertation constitute a useful tools for practical applications. Although they overcome most drawbacks of the classical W-F type procedures, it has to be pointed out that computational cost involved by their use may be still quite high, especially when considering high-dimensional spaces and multi-parameter models (e.g., large-scale 3D models of air pollution). The main reason behind this is the necessity of multiply solving the state equations and applying global optimization to ensure a reasonable quality of the resulting solutions. Thus, a natural trend to increase the performance is to exploit the benefits provided by parallel processing techniques. The remainder of the section briefly delineates the possibilities of using cluster computing in the search for T-optimum designs.

5.4.1. Message Passing Interface (MPI)

Modern scientific and engineering computations base on large-scale simulations. To achieve a reasonable performance, parallel environments which offer possibilities of computing on several processors at the same time are used. These environments often are called supercomputers.

A supercomputer is a computer that is powerful in terms of processing possibilities, particularly speed of calculations. The notion of supercomputer itself is rather vague since the persistent speedup in potential computational power yields depreciation of particular solutions. In the 1970s most supercomputers were dedicated to running vector processors. In the late 1980s and 1990s, attention turned to massive parallel processing systems with thousands of simple CPUs, some being off the shelf units and others being custom designs. Today, parallel designs base on 'off the shelf' RISC microprocessors, such as PowerPC or PA-RISC, and most modern supercomputers constitute highly-tuned computer clusters using commodity processors combined with custom interconnects (Landau *et al.*, 1993; Buyya, 1999).

There exist two main architectures of supercomputers to achieve parallel computing:

- Symmetric Multi-Processing (SMP) should be understood as a single computer with more then one processor. Multiple CPUs residing in one cabinet share the same memory. Generally, SMP are very expensive and are employed for specialized applications that require enormous amounts of floating-point computations. Such systems provide a rather small scalability. As more requirements are established, additional CPUs can be added, but there exist hardware, e.g., motherboard, constraints. The SMP systems range from two up to as many as hundreds or more processors. However, if one CPU fails, the entire SMP system is also useless. To prevent such a situation, clusters of two or more SMP systems can be used to provide high availability (fault resilience). If one SMP system fails, the others continue operations.

- Clusters. This architecture usually consists of many ordinary relatively low-cost computers (compared with SMP constructions) called computing nodes.

> In other words, a computer cluster is a group of loosely coupled computers that work together closely so that in many respects it can be viewed as if it were a single computer. Clusters are commonly connected through fast local area networks (GigaBit Ethernet, Myrinet, Infiniband) and are typically much more cost-effective. One of the open source ideas which support building computer clusters is Beowulf. Beowulf clusters are scalable performance clusters based on commodity hardware, on a private system network, with open source software (Linux) infrastructure. Clusters are highly scalable. The designer can improve performance proportionally with added machines. The commodity hardware can be a number of mass-market, stand-alone computing nodes as simple as two networked computers, each running Linux and sharing a file system, or as complex as 1024 nodes with a high-speed, low-latency network.
>
> Class I clusters are built entirely using commodity hardware and software using a standard technology such as SCSI, Ethernet, and IDE. They are typically less expensive than Class II clusters which may use specialized hardware to achieve higher performance. Clusters which consist of computing nodes that have the same hardware are called homogenous. Clusters with distinct architectures are called heterogenous clusters or heterogenous systems (Bookman, 2003).

The message passing model defines a set of processes that have access only to local memory but are able to communicate with other processes by sending and receiving messages. The major goal of Message Passing Interface (MPI), as with most standards, is to ensure portability across different machines (Pacheco, 1997; Snir *et al.*, 1998; Gropp *et al.*, 1998) which would be comparable to that of programming languages such as Fortran. This means that the same message passing source code can be executed on a variety of machines as long as an MPI library is available, while some tuning might be needed to take advantage of the features of each system. Though message passing is often defined in the context of distributed-memory parallel computers, the same code can run well on a shared-memory parallel computer. It can be run on a network of workstations or as a set of processes running on a single workstation. Knowing that efficient MPI standard implementations exist across a wide variety of computers gives a high degree of flexibility in the code development, debugging, and in choosing a particular platform for program execution.

Message Passing Interface provides possibilities of data transfer from the local memory of one process to the local memory of another process. Such a situation appears when data processing requires operations being performed by several distinct processes. There are many specific communication networks (Fast/GigaBit Ethernet, Infiniband, Myrinet) with different connection topologies.

In practice, the MPI is a library which provides a set of communication procedures for both parallel computers and workstation networks, especially dedicated for Unix/Linux clusters.

5.4.1.1. Basic MPI subroutines and program structure

In the message passing model of parallel computation, the processes executed in parallel have separate address spaces. Communication occurs when a portion of the address space of one process is copied into the address space of another process. This operation needs cooperation and occurs only when the first process executes a send operation and the second a receive operation, respectively. MPI provides a set of send and receive functions that allow for the transmission of data of a particular type. Additionally, the so-called tag is associated with the message. Typing of the message contents is necessary for heterogeneous environments — the type information is needed so that conversion of a correct data representation can be performed when data are sent from one architecture to another. The tag mechanisms provide the tool which allows the message being received selectively, i.e., at the particular receiving end only a message with a specified, particular tag can be received or a 'wild-card' value for this quantity can be used, allowing reception of messages with any tag.

An MPI library consists of C or Fortran 77 subroutines which allow communications between MPI processes by calling MPI routines. All names of MPI routines and constants in both C and Fortran begin with the prefix `MPI_` to avoid name collisions. Fortran routine names are all upper case. The following basic instructions are indispensable in Fortran implementations. An MPI program requires the use of the module MPI and begins with the include instruction `include 'mpif.h'`. This include file or module is necessary in every MPI Fortran program and subprogram to define various constants and variables. Then MPI subroutines are executed after the `MPI_INIT` and before `MPI_FINALIZE` routine calls. The call to `MPI_INIT` is required in every MPI program and must be the first MPI call. It establishes the MPI 'environment'. Only one invocation of this routine can occur in each program execution. The routine requires only an argument which is an error code. Every Fortran MPI subroutine returns an error code as its last argument which can take either value `MPI_SUCCESS` or an implementation-defined error code. The MPI consists of over a hundred subroutines, but for a simple parallel program it is sufficient to use only six basic procedures which are listed below (MacDonald *et al.*, 2005; Gropp *et al.*, 1999):

- `MPI_INIT(ierr)`

  Initialize the MPI enviroment.

- `MPI_COMM_SIZE(MPI_COMM_WORLD,size,ierr)`

  Determine the size of the group associated with the generally used communicator
  `MPI_COMM_WORLD`, which is a default communicator predefined in `mpif.h` and defines the scope of all processes (processors).

- `MPI_COM_RANK (MPI_COMM_WORLD, myrank, ierr)`

  Determine the rank of a process in the communicator. Initially, each process will be assigned a unique integer rank between zero and the number of

processors minus one within the communicator `MPI_COMM_WORLD`. This rank is often referred to as a task ID. If a process becomes associated with other communicators, it will have a unique rank within each of these as well.

- `MPI_SEND(buf, count, datatype, dest, tag, MPI_COMM_WORLD, ierr)`

  Basic blocking send operation. The routine terminates only after the application buffer in the sending task is free for reuse. Some implementations may actually use a synchronous send to implement the basic blocking send. The arguments are:

`buf`	address of the send buffer
`count`	number of items to send
`datatype`	data type of each item
`dest`	rank of destination process
`tag`	message tag

- `MPI_RECV(buf, count, datatype, source, tag, MPI_COMM_WORLD, status, ierr)`

  Receive a message and block until the requested data are available in the application buffer of the receiving task. The arguments are:

`buf`	initial address of the receive buffer
`count`	maximum number of elements in the receive buffer
`datatype`	data type of each receive buffer element
`dest`	rank of the source process
`tag`	message tag
`status`	return status

- `MPI_FINALIZE (...)`

  Terminate the MPI execution environment.

As an example of a very simple MPI program consider (MacDonald *et al.*, 2005):

```
PROGRAM test
       INCLUDE 'mpif.h'
       INTEGER myrank, size, ierr
       call MPI_INIT(ierr)
       call MPI_COMM_RANK(MPI_COMM_WORLD, myrank, ierr)
       call MPI_COMM_SIZE(MPI_COMM_WORLD, size, ierr)
       print *, "Processor", myrank, "of", size, ": MPI test"
       call MPI_FINALIZE(ierr)
END
```

which compiled (the compilation manner and commands depend a particular configuration of the environment, e.g., `mpif90 test.f90 -o test.out`) and invoked `mpirun -np 6 test.out` where the parameter `-np x` specifies the number of processors participating in computations, produces the following output:

```
Processor           0 of           6 : MPI test
```

```
Processor          1 of          6 : MPI test
Processor          2 of          6 : MPI test
Processor          3 of          6 : MPI test
Processor          4 of          6 : MPI test
Processor          5 of          6 : MPI test
```

5.4.2. Cluster computations of T-optimum designs

Since the most time-consuming part of the computational methods presented heretofore is the involved global optimization, it is natural that main attention was concentrated on parallelizing that part of computations. As a global optimizer, the stochastic ARS strategy described in Section 4.4 was used. The nature of stochastic methods results in a rather straightforward development of their parallel counterparts. Thus, in our approach a number of *slave* processors compute global minima or maxima in parallel and the reminder of the algorithm is executed by the *master* processor. Such a *master-slave* model is typical for MPI-based parallel programming (Gropp *et al.*, 1999).

A parallel implementation of the ARS scheme can be achieved using the following techniques (Kuczewski *et al.*, 2006):

❶ 'multistart', when each of the processors performs a search in the entire admissible domain, but using different initial guesses; additionally, each processor can perform several search cycles, all with different initial guesses;

❷ a partition of the search domain, when each processor conducts a search only in a proper subdomain of the entire set of feasible solutions; several search cycles per processor are also possible;

❸ a combination of the foregoing two methods.

Example 5.2. In order to verify the potential possibilities of reducing the total computational time, the problem of discrimination between two models of chemical reactions, similar to the one from Example 4.2 was considered (with slightly changed parameter values). To obtain a solution, the Wynn-Fedorov scheme described in Section 4.1 was used. We were mostly interested in a time speedup possible to achieve when using different numbers of processors.

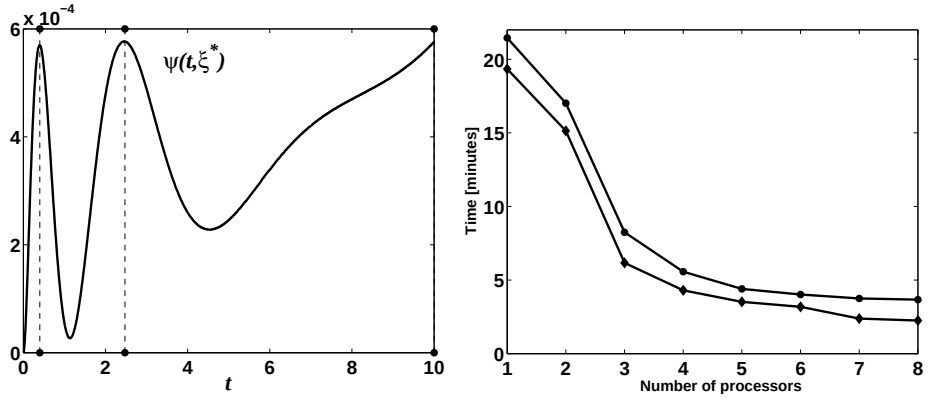
The program used for parallel computing of optimum designs was written completely in Fortran 95 using `ifort` (Intel®Fortran Compiler v.8.1 for Linux 64-bit platforms) and the open-source `mpich-1.2.6` implementation of MPI. Computations were performed on a Linux cluster which has been recently built at the University of Zielona Góra within the framework of the national CLUSTERIX project (Wyrzykowski *et al.*, 2004). This homogenous cluster is equipped with four SMP nodes, each node with two 64-bit Intel Itanium II 1.4GHz processors and 2GB RAM memory, running under the control of GNU/Linux Debian for ia64. Connection between the nodes is realized via Gigabit Ethernet.

The parallel ARS procedure was implemented using strategy ❶ ('multistart') and ❷ (domain partitioning) in the above mentioned *master-slave* model, when

one of the processors plays a managing role and assigns subtasks to the remaining processors, gathers partial results from them and then determines the ultimate solution in the sequential part of the program. In order to examine the time speedup, we performed numerous experimental runs and obtained virtually the same solution (optimum design) each time – differences resulting from numerical inaccuracies appeared only from the second decimal place. The resulting optimum design contains three support points and has the form

$$\xi^\star = \begin{Bmatrix} 0.40, & 2.47, & 10.00 \\ 0.229, & 0.327, & 0.444 \end{Bmatrix}. \tag{5.82}$$

This means that almost half of the measurement effort should be concentrated at the time instant $t = 10\,s$. As can be seen from Fig. 5.7(a) containing a plot of the sensitivity function $\psi(t) = \|\eta(t) - \eta_2(t, \vartheta_2^\star)\|^2$, the support points are located at the maxima of $\psi(t, \xi^\star)$, which is consistent with the presented T-optimality conditions.



(a) (b)

Fig. 5.7. Plot of the sensitivity function vs time and the location of the support points in the optimal design (vertical dashed lines) (a), and the plot of the total computation time as a function of the number of processors for strategies ❶ and ❷ (marked with ♦ and •, respectively) (b) in Example 5.2.

The compiled program was executed using various numbers of processors. The number of search cycles performed on each *slave* processor depended on the total number of processors to achieve altogether about 32 independent starts of the ARS algorithm per iteration of the Wynn-Fedorov procedure. Figure 5.7(b) shows the averaged total processing time as a function of the number of processors in the case of using both the strategies.

When analyzing the obtained speedup, it is worth of noticing that despite a constant, deterministic computation time of a single ARS run (since the number of generated trial points is fixed *a priori* and remains constant), the ARS algorithm itself is probabilistic. This means, that the quality of the obtained solution may

vary from run to run, which, in consequence, may influence the convergence of the Wynn-Fedorov scheme.

The trend and shape of the speedup curves is typical compared with the results obtained for parallelization of problems of stochastic search, e.g., genetic algorithms (Kwedlo, 2004). We have a noticeable difference in processing time between two and three processors. A further increase in the number of processors does not yield so strongly a decrease in the total computational time. Nevertheless, results are promising, especially when we consider the fact that parallelization was applied only in the phase of global optimization. Satisfactory modelling of complex spatial processes, like atmospheric or groundwater pollution proliferation leads to problems of discriminating between models of distributed parameter systems described by systems of PDE's. In that case, the level of computational complexity is much higher owing to the necessity of multiply solving model equations on relatively dense, multidimensional grids. It seems that combining parallelization of solving PDE's describing models with parallelization of the global optimization process should allow for a wider application of discrimination techniques based on optimum experimental design in fields such as pollution forecast or diagnosis of complex industrial processes.

5.5. Summary

Four practical problems associated with the behaviour of the optimization process when determining T-optimal designs were treated in some detail. Of fundamental importance was the study of the convergence of the Wynn-Fedorov algorithm for the newly proposed DT-optimality criterion. This constitutes the first result of this type, and our detailed proof revealed salient features of this scheme when applied to maximin criteria. Consequently, the strong assumptions on the structure of the answering sets (they have to be singletons) suggest that serious problems with convergence may be observed when trying to cope with complex models encountered in engineering practice. But then the RATO algorithm of Section 4.3 can be used, since the appropriate generalization of this procedure to handle the DT-optimality criterion is rather straightforward.

The other three issues discussed in this chapter were related to a possible reduction in the amount of time required to determine T-optimal designs. First, an important asset of the design procedure on finite design spaces, which was taken here, is that direct computational methods suggested themselves; that is, the computational algorithm of the gradient-projection type and the duality-based Uzawa method arose quite naturally in this context. Of significant importance is the fact that these algorithms rely on the concave nature of the T-optimality criterion treated as a function of the design weights. The resulting procedure can be included as a component in the RATO or Wynn-Fedorov algorithms, thereby considerably accelerating the computational process.

Another issue was the implementation of Rafajłowicz's selective random search technique based on application of a Markov chain Monte Carlo method. The MCMC sampling strategy is very easy to implement and broadly applicable. Ac-

tive interest in this area in recent years, which has changed a significant part of the everyday practice of statistics, and promising results reported here suggest that this line of investigations should be one of the main directions for future research.

Finally, a way to parallelize the Wynn-Fedorov algorithm has been presented. The most time-consuming subprocedure, i.e., global optimization, has been parallelized, and this has been done within the MPI (Message Passing Interface) framework. Decreases in computational times can be significant, although because of the simplicity of the exemplary problem the results reported do depend on the communication between nodes. Gains are expected to be even higher for more complex problems if more sophisticated parallelization strategies are employed. This will constitute another prospective research direction in the near future.

Chapter 6

DESIGNS WITH APPLICATION-DRIVEN OBJECTIVES

The aim of this chapter is to develop some systematic approaches to T-optimum experimental design oriented towards applications. Thus, a more realistic problem of discrimination between several, instead of only two, competing models is delineated. Then, in Section 6.2 the notion of the so-called replication-free designs is used to derive an extremely useful approach which is free from the necessity of taking multiple measurements at one design point. In turn, it leads to simple and efficient exchange algorithm, allowing us to locate a given number of measurements on a finite set of discrete admissible locations, e.g, a grid of possible locations. Such a situation is going to be common in engineering practice, e.g., in environmental research where we deal with problems of monitoring air or water quality using monitoring sensor networks. This type of problems is strictly connected with the necessity of modelling complex spatio-temporal processes. The phenomena like pollutant proliferation in the atmosphere require descriptions in the form of partial differential equations. Thus, the application of T-optimum design to distributed parameter systems when placing stationary sensors is also considered in Section 6.3. Section 6.4 deals with an even more challenging problem of a design in the presence of correlated observation errors. To derive the appropriate design criterion, the so-called heteroscedastic framework is used. Finally, an application in the field of fault detection and diagnosis in industrial processes is considered in Section 6.5.

6.1. Discrimination between several dynamic models

The considerations presented in this section constitute a straightforward generalization of the problem of determining an optimum design for discrimination between only two models. In the sequel we assume that there exist v candidate models. We shall concentrate on general non-linear dynamic models described by the following ordinary differential equations:

$$\mathcal{M}_\ell : \quad \frac{\mathrm{d}\eta_\ell(t)}{\mathrm{d}t} = f_\ell(t, \eta_\ell(t), \widetilde{\vartheta}_\ell), \quad \eta_\ell(0) = \eta_0, \quad \ell = 1, \ldots, v, \tag{6.1}$$

where t is time, η stands for a vector-valued function $\eta : T \to \mathbb{R}^d$ (also called the state), $T = [0, t_f]$ for a given t_f, $\widetilde{\vartheta}_\ell \in \mathbb{R}^{m_\ell}$ signifies a vector of constant

parameters being unknown to the experimenter, and $f_\ell : T \times \mathbb{R}^{d+m_\ell}$ is required to be continuous together with its Jacobi matrices $\partial f/\partial \eta_\ell$ and $\partial f/\partial \vartheta_\ell$, $\ell = 1, 2, \ldots, v$. Note that η_ℓ depends implicity on $\widetilde{\vartheta}_\ell$ (it will be stressed in the notation and further considerations as $\eta_\ell(t, \widetilde{\vartheta}_\ell)$).

Process responses can be observed up to additive random errors. Thus our basic assumption is that the discrimination between models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_v$ is to be performed by making observations $y_{ij} \in \mathbb{R}^d$ of the true process responses $\eta(t)$ at discrete time instants $t_1, \ldots, t_n$. The output equation is defined by the following regression model:

$$y_{ij} = \eta(t_i) + \varepsilon_{ij}, \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, n, \tag{6.2}$$

The nonlinear functional relationship $\eta : T \times \mathbb{R}^m \to \mathbb{R}^d$ constitutes the *true* model of the process. In our description, t_i stands for a setting of the independent variable $t \in T$ (the so-called *explanatory* or *regressor* variable, here the time instant of the measurement). The terms ε_{ij} represent random fluctuations (measurement errors) resulting, e.g., from inaccuracies in the measuring devices. The errors ε_{ij} are sampled from a distribution satisfying

$$E\big[\varepsilon_{ij}\big] = 0, \quad E\big[\varepsilon_{ij}\varepsilon_{\kappa\ell}^{\mathrm{T}}\big] = \begin{cases} \sigma^2 I_d & \text{if } i = \kappa \text{ and } j = \ell, \\ 0_d & \text{otherwise,} \end{cases} \tag{6.3}$$

I_d being the $d \times d$ identity matrix, 0_d the $(d \times d)$-dimensional matrix of zeros, and σ^2 a constant positive variance. The additional index j is necessary if the observations are to be repeated several times for some settings t_i, as in practice repeated experimental runs typically lead to different observed responses, even if t_i's are exactly the same. Here the number of replications for a given setting t_i is denoted by r_i, $\sum_{i=1}^n r_i = N$. The t_i's are called control variables because they can be chosen by the experimenter. They define process conditions and may vary from observation to observation (in (6.2) we thus have n different settings denoted by $t_1, \ldots, t_n$).

Like in previous chapters, our basic assumption is that the process response $\eta(t)$ coincides with one of the $\eta_\ell(t, \widetilde{\vartheta}_\ell)$'s, where functions $\eta_\ell(t, \widetilde{\vartheta}_\ell)$ are given *a priori*, with $\widetilde{\vartheta}_\ell \in \Theta_\ell \subset \mathbb{R}^{m_\ell}$ being constant parameters which are fixed but unknown to the experimenter except for the model which is assumed to be true (the Θ_ℓ's denote some known compact sets).

Without loss of generality we can again assume that the first model is true, i.e. $\eta(t) = \eta_1(t, \widetilde{\vartheta}_1)$ and that $\widetilde{\vartheta}_1$ is known to the experimenter. Then, in the case of v candidate models, we can define $v - 1$ non-centrality parameters:

$$\Delta_{1\ell}^0(\xi_N) = \min_{\vartheta_\ell \in \Theta_\ell} \sum_{i=1}^n p_i \|\eta(t_i) - \eta_\ell(t_i, \vartheta_\ell)\|^2, \quad \ell = 2, \ldots, v, \tag{6.4}$$

where

$$\xi_N = \begin{Bmatrix} t_1, & \ldots, & t_n \\ p_1, & \ldots, & p_n \end{Bmatrix}$$

constitutes the exact design and $p_i = r_i/N$, $i = 1, \ldots, n$. The estimates of the parameters $\widetilde{\vartheta}_\ell$, as before, are obtained using the least-squares method, i.e.,

$$\widehat{\vartheta}_{\ell N} = \arg \min_{\vartheta_\ell \in \Theta_\ell} \sum_{i=1}^{n} p_i \|y_i - \eta_\ell(t_i, \vartheta_j)\|^2, \quad \ell = 2, \ldots, v, \tag{6.5}$$

where

$$y_i = \frac{1}{r_i} \sum_{j=1}^{r_i} y_{ij}. \tag{6.6}$$

Recall that in the case of discriminating between only two models, the solution consists in maximizing the parameter $\Delta_1^0(\xi_N)$. Now the problem becomes that of discriminating between the true model and the one(s) closest to it. Therefore, following (Atkinson and Fedorov, 1975b), as the design criterion we take the maximum of the non-centrality parameters for one or more of the closest models and seek

$$\max_{\xi_N} \Delta_{(1)}^0(\xi_N), \tag{6.7}$$

where

$$\Delta_{(1)}^0(\xi_N) = \min_{2 \le \ell \le v} \Delta_{1\ell}^0(\xi_N). \tag{6.8}$$

In the case considered here, when we assume that only one closest model exists, this criterion constitutes a generalization of the one applied to a two-candidate case. A more complicated situation takes place when maximization of the minimum non-centrality parameter leads to designs for which there are two or more closest models (e.g., nested models) with equal non-centrality parameter values (Atkinson and Fedorov, 1975b).

To simplify the solution of the original discrete problem (6.7), we replace it by the continuous one

$$\max_{\xi} \Delta_{(1)}(\xi) \tag{6.9}$$

with

$$\Delta_{(1)}(\xi) = \min_{2 \le \ell \le v} \Delta_{1\ell}(\xi), \tag{6.10}$$

where ξ stands for any probability measure on the sigma-field of Borel subsets of the design space T. Again, the set of all such measures will be denoted by $\Xi(T)$.

Accordingly, as a continuous generalization of the non-centrality parameters, we consider the quantities

$$\Delta_{1\ell}(\xi) = \min_{\vartheta_\ell \in \Theta_\ell} \int_T \|\eta(t) - \eta_\ell(t, \vartheta_\ell)\|^2 \, \xi(\mathrm{d}t), \quad \ell = 2, \ldots, v. \tag{6.11}$$

The design

$$\xi^\star = \arg \max_{\xi \in \Xi(T)} \Delta_{(1)}(\xi) \tag{6.12}$$

will then be called the *local $T_{(1)}$-optimum design*. (The notation $T_{(1)}$ stresses the assumption that the first among rival models is assumed to be the true one.)

6.1.1. Optimality conditions

An easy generalization of the results presented in (Atkinson and Fedorov, 1975b) allows us to prove the following result, crucial for a construction of numerical approximations of optimum designs:

Let us make the following assumptions:

(C1) T and Θ_ℓ are compact sets for $\ell = 2, \ldots, v$,

(C2) $\eta(\,\cdot\,)$ is a continuous function on T,

(C3) $\eta_\ell(\,\cdot\,,\,\cdot\,)$ are continuous functions on $T \times \Theta_\ell, \quad \ell = 2, \ldots, v$.

Theorem 6.1. *Assume that there exists only one closest model, i.e., there exists a unique solution $\ell^\star$ of the subproblem (6.10) for a design $\xi^\star$. Moreover, suppose that the least-squares estimates of the parameters $\vartheta^\star_{\ell^\star} \in \Theta_{\ell^\star}$ are unique. Under Assumptions (C1)–(C3) a necessary and sufficient condition for $\xi^\star$ to be $T_{(1)}$-optimal is that, for every $t \in T$, we have*

$$\|\eta(t) - \eta_{\ell^\star}(t, \vartheta_{\ell^\star})\|^2 \le \Delta_{(1)}(\xi^\star). \tag{6.13}$$

The equality in (6.13) is attained at all support points of $\xi^\star$. Furthermore, the set of all the corresponding optimal measures $\xi^\star$ is convex.

6.1.2. Numerical construction of $\mathbf{T_{(1)}}$–optimum designs

The algorithm proposed below actually constitutes a modified version of Algorithm 4.1. Assuming that n^k stands for the number of support points in the current step, the algorithm can be represented by the following steps (Kuczewski, 2004):

Algorithm 6.1 (Discrimination between several models)**.**

Step 1: Guess an initial design ξ^0. Choose some positive tolerance $\epsilon \ll 1$. Set $k = 0$.

Step 2: In the k-th step, for each $\ell = 2, \ldots, v$, find

$$\widehat{\vartheta}^k_\ell = \min_{\vartheta_\ell \in \Theta_\ell} \sum_{i=1}^{n^k} p_i^k \|\eta(t_i^k) - \eta_\ell(t_i^k, \vartheta_\ell)\|^2. \tag{6.14}$$

Rank the residual sums of squares obtained for each model and determine the model producing the least value of the deviation from the true model:

$$\ell^k = \arg \min_{2 \le \ell \le v} \sum_{i=1}^{n^k} p_i^k \|\eta(t_i^k) - \eta_\ell(t_i^k, \widehat{\vartheta}^k_\ell)\|^2. \tag{6.15}$$

Then find

$$\hat{t}^k = \max_{t \in T} \|\eta(t) - \eta_{\ell^k}(t, \widehat{\vartheta}^k_{\ell^k})\|^2. \tag{6.16}$$

Step 3: If $\psi_{1\ell}(\hat{t}^k, \xi^k) \leq \Delta_{1\ell}(\xi^k) + \epsilon$, where

$$\psi_{1\ell}(t, \xi^k) = \|\eta(t) - \eta_{\ell^k}(t, \widehat{\vartheta}^k_{\ell^k})\|^2, \tag{6.17}$$

$$\Delta_{1\ell}(\xi^k) = \sum_{i=1}^{n^k} p^k_i \|\eta(t^k_i) - \eta_{\ell^k}(t^k_i, \widehat{\vartheta}^k_{\ell^k})\|^2, \tag{6.18}$$

then set $\xi^\star = \xi^k$ and STOP.

Step 4: Choose an appropriate α^k with $0 \leq \alpha^k \leq 1$ and compute the convex combination of designs:

$$\xi^{k+1} = (1 - \alpha^k)\xi^k + \alpha^k \delta(\hat{t}^k) \tag{6.19}$$

where $\delta(\hat{t}^k)$ is the unit-weight design concentrated at the new trial point $\hat{t}^k$. Set $k \leftarrow k + 1$ and go to Step 2.

The above procedure consists in taking a new trial point at which the discrepancy between the responses of the true and best-fit models are furthest apart. Since it constitutes only some modification of the generalized Wynn-Fedorov scheme, its properties and computational requirements are very similar (the latter are increased owing to the necessity of computing $v - 1$ times least-squares solutions $\widehat{\vartheta}^k_\ell$ during Step 2).

Example 6.1. Consider the process of chemical conversion of substance A into B and C in a batch reactor (Stewart *et al.*, 1998). Four rival models are considered for this system:

Model 1: Consecutive reversible first-order reactions $A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \underset{k_4}{\overset{k_2}{\rightleftharpoons}} C$ described by the following equations with the appropriate initial conditions:

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A] + k_3[B], \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A] - (k_2 + k_3)[B] + k_4[C], \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B] - k_4[C], \\ [A]_{t=0} = a_0, \quad [B]_{t=0} &= b_0, \quad [C]_{t=0} = c_0, \end{aligned} \tag{6.20}$$

where $[A]$, $[B]$ and $[C]$ denote the concentrations of the reagents A, B and C, respectively.

Model 2: Consecutive irreversible first-order reactions $A \overset{k_1}{\rightarrow} B \overset{k_2}{\rightarrow} C$ described by

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A], \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A] - k_2[B], \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B], \\ [A]_{t=0} = a_0, \quad [B]_{t=0} &= b_0, \quad [C]_{t=0} = c_0. \end{aligned} \tag{6.21}$$

Model 3: A reversible first-order reaction followed by an irreversible first-order reaction
$A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \overset{k_2}{\to} C$ described by

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A] + k_3[B], \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A] - (k_2 + k_3)[B], \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B], \\ [A]_{t=0} = a_0, \quad [B]_{t=0} &= b_0, \quad [C]_{t=0} = c_0. \end{aligned} \tag{6.22}$$

Model 4: It is formed out of Model 2 supplemented by an irreversible first-order reaction $A \overset{k_3}{\to} C$, which gives

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A], \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= k_1[A] - k_2[B], \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_2[B] + k_3[A], \\ [A]_{t=0} = a_0, \quad [B]_{t=0} &= b_0, \quad [C]_{t=0} = c_0. \end{aligned} \tag{6.23}$$

In the above models, parameters k_i, called the rates, are responsible for the dynamics of concentrations changes. Model 1 is the true one, since it was used to generate the data with parameters $\widetilde{\vartheta}_1 = (k_1^1, k_2^1, k_3^1, k_4^1) = (0.7, 0.7, 0.1, 0.1)$.

The aim of the design is to determine an optimum schedule of the concentration measurements, which would allow us to answer the question if the considered reaction can be described by one of the alternative, structurally simpler models (accepting some unavoidable, small level of discrepancy resulting from different structures).

The admissible ranges of the parameters k_i^ℓ for the ℓ-th rival model were fixed as $0.3 \le k_1^2, k_2^2, k_1^3, k_2^3, k_1^4, k_2^4 \le 1.0$, $0.1 \le k_3^3, k_3^4 \le 0.5$. The time horizon was set as $T = [0, 10]$ and the initial concentrations were fixed as $(a_0, b_0, c_0) = (1.0, 0.0, 0.0)$. The code to calculate the approximation of the optimum design using Algorithm 6.1 was run using the Lahey-Fujitsu Fortran 95 compiler v5.6 with IMSL library. After 44 iterations of the procedure, Model 3 turned out to be the best fit among the three candidates (the value of the non-centrality parameter was $\Delta_{13} = 0.0106$). Model 2 was slightly worse fit ($\Delta_{12} = 0.0117$) but still being well ahead of Model 4 with $\Delta_{14} = 0.2965$. The obtained optimum design is of the form

$$\xi^\star = \begin{Bmatrix} 0.77, & 3.45, & 10.00 \\ 0.0952, & 0.3095, & 0.5952 \end{Bmatrix}. \tag{6.24}$$

The least profitable parameters of the best fit model (Model 3) compared with the optimum design $\xi^\star$ amount to $\vartheta_3^\star = (k_1^3, k_2^3, k_3^3) = (0.9609,\ 0.3708,\ 0.3756)$.

The sensitivity function $\psi_{13}(t, \xi^\star) = \|\eta(t) - \eta_3(t, \vartheta_3^\star)\|^2$, defining discrepancies in the responses of the true and best fit models has the form shown in Fig. 6.1. It is worth of noticing that the support points are located in the places where the sensitivity function attains its upper bound Δ_{13}, which is consistent with the theory developed and confirms the optimality of the design. The responses of the true and two successively best fit alternative models are shown in Fig. 6.2.

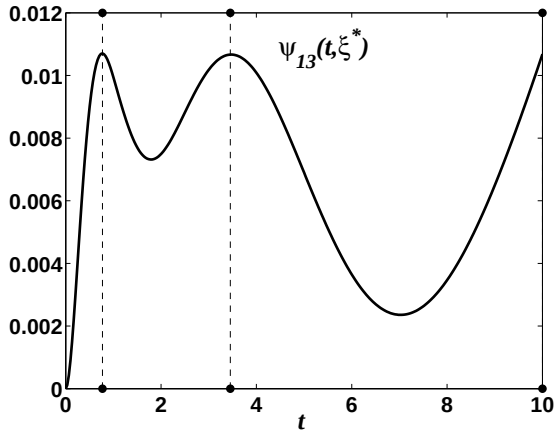


Fig. 6.1. Sensitivity function $\psi_{13}(t, \xi^\star)$ for the best fit model (Model 3) and the location of the support points (dashed vertical lines) in Example 6.1.

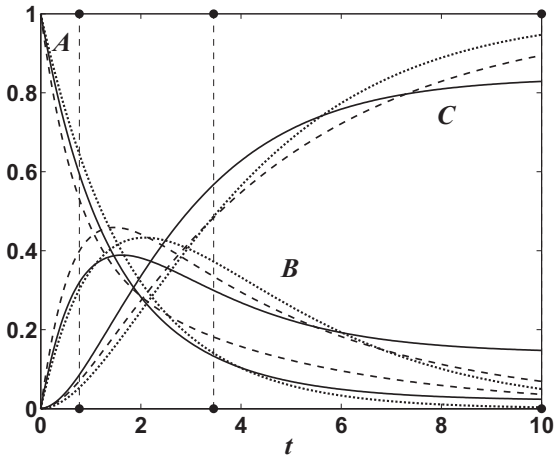


Fig. 6.2. Responses of the true, third and second models (solid, dashed and dotted lines, respectively) vs time in Example 6.1. The support points of the optimal design are also indicated (dashed vertical lines).

6.2. Replication-free T-optimum designs

In the existing literature, as well as in the material presented so far, an underlying assumption on T-optimal designs is that measurements at optimal time instants or in optimal places can be repeated several times if necessary. In many situations, however, this condition is too restrictive as such replications are impossible for practical reasons (e.g., only one run of the considered process is allowed or repetitions are simply too expensive to perform). The experimental design consists then in an adequate choice of observation instants or places. Thus in the present section, a replication-free measurement policy with a fixed number of measurement times or points is proposed. For that purpose, the idea of operating on the density of measurements, rather (Fedorov, 1989; Cook and Fedorov, 1995; Fedorov and Hackl, 1997) that on the measurement instances is exploited.

The considerations below are presented for dynamic systems such that the continuous generalization of the T-optimality criterion has the form

$$\Delta_1(\xi) = \min_{\vartheta_2 \in \Theta_2} \int_T \|\eta(t) - \eta_2(t, \vartheta_2)\|^2 \, \xi(\mathrm{d}t). \tag{6.25}$$

In the reminder of the section, we shall make the assumption that the problem of minimizing the performance index $S_\xi(\vartheta_2) = \int_T \|\eta(t) - \eta_2(t, \vartheta_2)\|^2 \, \xi(\mathrm{d}t)$ with respect to ϑ_2 possesses a unique solution $\widehat{\vartheta}_2(\xi)$ for any fixed ξ. Then we can prove that

$$\Delta_1((1-\alpha)\xi + \alpha\mu) = \Delta_1(\xi) + \alpha \int_T \psi(t, \xi) \, \mu(\mathrm{d}t) + o(\alpha; \xi, \mu), \tag{6.26}$$

for any $\xi, \mu \in \Xi(T)$, where

$$\begin{aligned} &\lim_{\alpha \downarrow 0} o(\alpha; \xi, \mu)/\alpha = 0, \\ &\psi(t, \xi) = \|\eta(t) - \eta_2(t, \widehat{\vartheta}_2(\xi))\|^2 - \int_T \|\eta(t) - \eta_2(t, \widehat{\vartheta}_2(\xi))\|^2 \, \mu(\mathrm{d}t). \end{aligned} \tag{6.27}$$

In order to avoid replicated measurement configurations, we implement the idea of operating on the density of measurements (i.e., the number of measurements per unit time interval), rather than on the measurement locations, which is justified for a sufficiently large total number of measurements N. In contrast to the classical designs, however, we impose the crucial restriction that the density of measurements must not exceed some prescribed level. For a design measure $\xi(\mathrm{d}t)$ this amounts to the condition

$$\xi(\mathrm{d}t) \leq \omega(\mathrm{d}t), \tag{6.28}$$

where $\omega(\mathrm{d}t)$ signifies the maximal possible 'number' of measurement per $\mathrm{d}t$ (Fedorov and Hackl, 1997) such that

$$\int_T \omega(\mathrm{d}t) \geq 1. \tag{6.29}$$

Consequently, we are faced with the following optimization problem: Find

$$\xi^\star = \arg \max_{\xi \in \Xi(T)} \Delta_1(\xi) \tag{6.30}$$

subject to

$$\xi(\mathrm{d}t) \le \omega(\mathrm{d}t). \tag{6.31}$$

The design $\xi^\star$ above is then said to be a (T, ω)-*optimal design* (Fedorov and Hackl, 1997; Uciński, 2005).

Apart from the usual assumptions made for T-optimum designs, a proper mathematical formulation calls for the proviso that $\omega(\mathrm{d}t)$ is atomless, i.e., for any $\Delta T \subset T$ there exists a $\Delta T' \subset \Delta T$ such that

$$\int_{\Delta T'} \omega(\mathrm{d}t) < \int_{\Delta T} \omega(\mathrm{d}t). \tag{6.32}$$

In what follows, we write $\Xi(\bar{T})$ for the collection of all the design measures which satisfy the requirement

$$\xi(\Delta T) = \begin{cases} \omega(\Delta T) & \text{for } \Delta T \subset \operatorname{supp} \xi, \\ 0 & \text{for } \Delta T \subset T \setminus \operatorname{supp} \xi. \end{cases} \tag{6.33}$$

Given a design ξ, we will say that the function $\psi(\,\cdot\,, \xi)$ defined by (6.27) separates sets T_1 and T_2 with respect to $\omega(\mathrm{d}t)$ if for any two sets $\Delta T_1 \subset T_1$ and $\Delta T_2 \subset T_2$ with equal non-zero ω-measures we have

$$\int_{\Delta T_1} \psi(t, \xi)\, \omega(\mathrm{d}t) \ge \int_{\Delta T_2} \psi(t, \xi)\, \omega(\mathrm{d}t). \tag{6.34}$$

We can now formulate the main result which provides a characterization of (T, ω)-optimal designs.

Theorem 6.2.

(i) There exists an optimal design $\xi^\star \in \Xi(\bar{T})$, and

(ii) A necessary and sufficient condition for $\xi^\star \in \Xi(\bar{T})$ to be (T, ω)-optimal is that $\psi(\,\cdot\,, \xi^\star)$ separates $T^\star = \operatorname{supp} \xi^\star$ and its complement $T \setminus T^\star$ with respect to the measure $\omega(\mathrm{d}t)$.

Proof. The result may be proved in much the same way as Theorem 4.3.1 of (Fedorov and Hackl, 1997, p. 63), also see (Cook and Fedorov, 1995) or (Uciński, 2005). □

From a practical point of view, the above theorem means that at all the support points of an optimal design $\xi^\star$ the mapping $\psi(\,\cdot\,, \xi^\star)$ should be greater than anywhere else, i.e., preferably $\operatorname{supp} \xi^\star$ should coincide with maximum points

of $\psi(\,\cdot\,,\xi^\star)$ If we were able to construct a design with this property, then it would be qualified as an optimal design. This conclusion forms a basis for numerical algorithms of constructing solutions to the problem under consideration.

As regards the interpretation of the resultant optimal designs (provided that we are in a position to calculate at least their approximations), one possibility is to partition T into subdomains ΔT_i of relatively small areas and then to allocate to each of them the number

$$N^\star(\Delta T_i) = \left\lceil N \int_{\Delta T_i} \xi^\star(\mathrm{d}t) \right\rceil \tag{6.35}$$

of measurements whose positions may coincide with nodes of some uniform grid (Fedorov and Hackl, 1997; Uciński, 2005) (here $\lceil\zeta\rceil$ is the smallest integer greater than or equal to ζ). Additionally, bear in mind that we must also have $\xi^\star(\mathrm{d}t) = \omega(\mathrm{d}t)$ in $T^\star$.

Clearly, unless the considered design problem is quite simple, we must employ a numerical algorithm to make the outlined conceptions useful. Since $\xi^\star(\mathrm{d}t)$ should be non-zero in the areas where $\psi(\,\cdot\,,\xi^\star)$ takes on a larger value, the central idea is to move some measure from areas with smaller values of $\psi(\,\cdot\,,\xi^k)$ to those with larger values, as we expect that such a procedure will improve ξ^k. This forms a basis for the numerical algorithm.

6.2.1. Replication-free numerical scheme

In what follows, we propose an exchange algorithm for computing replication-free designs. The name follows from the property that we sequentially perform replication-free exchanges of support points (a particular point may appear only once in the design and only one observation can be performed at each point). The number of support points n is thus constant and chosen *a priori* before the experiment. Assume that a sufficiently dense grid of discrete points $T_d = \{\tau_1, \tau_2, \ldots, \tau_R\} \subset T$ at which observations of the system output can be taken is given. The algorithm can be represented by the following steps (Kuczewski, 2003) (note that all design weights are now equal to $1/n$):

Algorithm 6.2 (Replication-free algorithm for T-optimum designs)**.**

Step 1: Guess an initial n-point replication-free design ξ^0. Choose some positive tolerance $\epsilon \ll 1$. Set $k = 0$.

Step 2: Find

$$\widehat{\vartheta}_2^k = \arg\min_{\vartheta_2\in\Theta_2} \sum_{i=1}^{n} \|\eta(t_i^k) - \eta_2(t_i^k, \vartheta_2)\|^2. \tag{6.36}$$

Determine

$$\begin{aligned} t_1^k &= \arg\min_{t\in T_1^k} \|\eta(t) - \eta_2(t, \widehat{\vartheta}_2^k)\|^2, \\ t_2^k &= \arg\max_{t\in T_2^k} \|\eta(t) - \eta_2(t, \widehat{\vartheta}_2^k)\|^2, \end{aligned} \tag{6.37}$$

where $T_1^k = \text{supp}(\xi_k)$ stands for the set of support points at stage k, $T_2^k = T_d \setminus T_1^k$.

Step 3: If $\psi(t_2^k, \xi^k) \leq \psi(t_1^k, \xi^k) + \epsilon$, where

$$\psi(t, \xi^k) = \|\eta(t) - \eta_2(t, \widehat{\vartheta}_2^k)\|^2, \tag{6.38}$$

then $\xi^\star = \xi^k$ and STOP.

Step 4: Swap t_1^k and t_2^k (exchange without replication), i.e., set $T_1^{k+1} = (T_1^k \setminus \{t_1^k\}) \cup \{t_2^k\}$, update ξ^{k+1} accordingly, set $k \leftarrow k+1$ and go to Step 2.

The basic concept of the procedure is very intuitive since we should obtain a design including points of maximum discrepancy between models. Computationally, Step 2 is of crucial significance but at the same time it is the most time-consuming part of the procedure. Complications arise, among other things, due to the necessity of calculating a global optimum in (6.36) (getting stuck in one of local optima may result in a precocious termination of the algorithm). Therefore, while implementing this part of the procedure, an effective global optimizer is essential, just as was the case in previous chapters. Determination of a pair t_1^k and t_2^k is definitely easier, since it reduces to a simple search on finite sets T_1^k and T_2^k, respectively.

Example 6.2. To illustrate the performance of the algorithm, the models from Example 4.3 were used, but with slightly changed parameters. The true model, describing pendulum movements has the form

$$\begin{aligned} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} &= \omega(t), & \varphi(0) &= \pi/2, \\ \frac{\mathrm{d}\omega(t)}{\mathrm{d}t} &= -10\sin(\varphi(t)), & \omega(0) &= 0, \end{aligned} \tag{6.39}$$

where $\varphi(t)$ and $\omega(t)$ stand for the angle and the angular velocity, respectively. The rival model has the structure

$$\begin{aligned} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} &= \omega(t), & \varphi(0) &= \pi/2, \\ \frac{\mathrm{d}\omega(t)}{\mathrm{d}t} &= -\vartheta_2\varphi(t), & \omega(0) &= 0, \end{aligned} \tag{6.40}$$

where ϑ_2 is an unknown parameter.

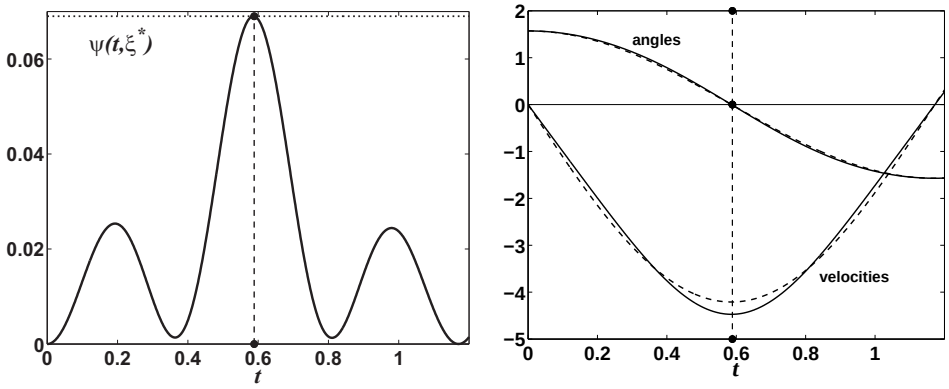
In order to compare properties of the classical and replication-free designs, we present results of using both the standard approach with replication allowed (using the Wynn-Fedorov scheme) and the clusterization-free approach.

The design horizon was fixed as $T = [0,\ 1.2]$ and the set of admissible parameter values was $\Theta_2 = [5,\ 10]$ in the case of both the experiments. The programs to calculate optimum designs were written in the Lahey-Fujitsu Fortran 95 v5.6

environment using some procedures from the IMSL library. The resultant classical design includes only one support point at which the whole measurement effort should be concentrated:

$$\xi^\star = \begin{Bmatrix} 0.5887 \\ 1.0 \end{Bmatrix}. \tag{6.41}$$

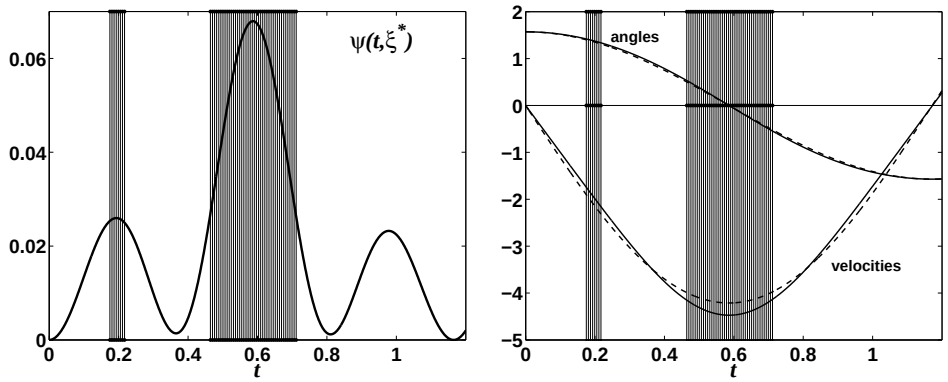
The least profitable parameter of the alternative model with respect to the optimum design $\xi^\star$ is $\vartheta_2^\star = 7.1813$. The support point coincides with the maximum point of the sensitivity function $\psi(t, \xi^\star) = \|\eta(t) - \eta_2(t, \vartheta_2^\star)\|^2$, which stands for the point of the maximum discrepancy between the responses of both the models (Fig. 6.3).



(a) (b)

Fig. 6.3 . Sensitivity function (a), and true and alternative model responses (solid and dashed lines, respectively) (b) obtained in Example 6.2 in the case of allowed replications, vs the location of the support point (the dashed vertical line).

During the second experiment, Algorithm 6.2 was employed to obtain an optimum replication-free design. It was assumed that the same time horizon was divided into 200 evenly spaced potential measurement points, from among which 50 were ultimately included in the design. The initial 50-point design ξ^0 was generated at random. The algorithm was run several times in order to avoid getting stuck in a local solution. The location of support points in the resulting optimal design is shown in Fig. 6.4. Note that the selected support points occupy the places where the discrepancies between the model responses in the case of their best fit are maximum. The optimal value of the second model parameter is almost identical to that in the previous experiment ($\vartheta_2^\star = 7.1884$). Small differences in these values and an asymmetric location of the support points in the resultant replication-free design are forced by inaccuracies in the numerical optimization.



(a) (b)

Fig. 6.4. Sensitivity function (a), and true and alternative model responses (solid and dashed lines, respectively) (b) obtained in Example 6.2 in the case of the T-optimal replication-free design, vs the location of the support points (vertical lines).

6.3. Discrimination between models of distributed parameter systems

When considering the systems with spatio-temporal dynamics described by partial differential equations, the optimal measurement problem appears (Uciński, 2005). This is related to the fact that there are several possible manners in which the measurements in spatio-temporal domains are made (Chen and Seinfeld, 1975). Measurements performed over the entire spatial domain continuously in time or at discrete time moments are considered to be of little significance, since it is generally not possible to carry out measurements over the whole spatial area Ω. Thus, measurements at discrete spatial locations are commonly encountered in engineering applications and, as a consequence, they dominate in the literature (Uciński, 2005; Rafajłowicz, 1978; Rafajłowicz, 1984; Rafajłowicz, 1986; Banks and Kunisch, 1989). They also can be performed continuously in time, or at discrete points in time.

If the measurements are not costly, then one would gather data as frequently as possible. As long as the timing of measurements is not a decision variable, the above-mentioned two manners are basically equivalent. Thus we focus here on the problem of T-optimum location of stationary sensors performing measurements continuously in time.

We now consider two competing models of distributed parameter systems defined in a spatial domain Ω and described by the following, possibly non-linear equations:

$$\mathcal{M}_\ell : \quad \frac{\partial \eta_\ell}{\partial t} = \mathcal{F}_\ell \left(x, t, \eta_\ell, \nabla \eta_\ell, \nabla^2 \eta_\ell, \vartheta_\ell \right), \quad \ell = 1, 2, \tag{6.42}$$

where $x \in \Omega$, $t \in T$ stands for time, $T = [0, t_f]$, $\eta_\ell = \eta_\ell(x, t, \vartheta_\ell)$ denotes the scalar state variable with values in $\mathbb{R}$ and $\mathcal{F}$ is some known function which may include terms accounting for forcing inputs given *a-priori*. Here $\vartheta_\ell \in \Theta_\ell$ denotes a vector of constant but unknown parameters, Θ_ℓ being a given compact set. For simplicity, we assume that both the equations are accompanied by the same boundary and initial conditions

$$\begin{cases} \mathcal{B}(x, t, \eta, \nabla\eta) = 0, & (x, t) \in \partial\Omega \times T, \\ \mathcal{P}(x, \eta, \nabla\eta) = 0, & (x, t) \in \Omega \times \{0\}, \end{cases} \tag{6.43}$$

where $\partial\Omega$ denotes a smooth boundary of Ω, $\mathcal{B}$ and $\mathcal{P}$ being known functions.

If we assume that the system is observed via N pointwise sensors placed at points $x^1, \ldots, x^n$, $n \leq N$ which take measurements at fixed time instants $t_1, \ldots, t_K \in T$, the output equation has the form

$$y_{ij}^k = \eta(x^i, t_k) + \varepsilon_{ij}, \quad j = 1, \ldots, r_i, \quad i = 1, \ldots, n, \quad k = 1, \ldots, K, \tag{6.44}$$

where $\eta(\,\cdot\,,\,\cdot\,)$ stands for the response of the true spatio-temporal system, r_i denotes the number of sensors (or, alternatively, replicated observations) at point x^i and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ represents a sequence of uncorrelated random measurement errors.

Now we directly follow the approach outlined in Chapter 3, i.e., we assume that the response of the first model is consistent with the true response of the process, i.e., $\eta(\,\cdot\,,\,\cdot\,) \equiv \eta_1(\,\cdot\,,\,\cdot\,,\widetilde{\vartheta}_1)$ for some known value of $\widetilde{\vartheta}_1 \in \Theta_1$.

Thus, the exact T-optimum design should maximize the value of the non-centrality parameter in the form

$$\Delta_1^0(\xi_N) = \min_{\vartheta_2 \in \Theta_2} \mathfrak{J}(\xi_N, \vartheta_2), \tag{6.45}$$

where

$$\mathfrak{J}(\xi_N, \vartheta_2) = \sum_{i=1}^{n} p_i \sum_{k=1}^{K} \|\eta(x^i, t_k) - \eta_2(x^i, t_k, \vartheta_2)\|^2, \tag{6.46}$$

and

$$p_i = r_i/N, \quad \sum_{i=1}^{n} p_i = 1.$$

Then we extend the notion of the design to the continuous case in much the same way as in Chapter 2. Moreover, for a sufficiently large K and the points $t_1, \ldots, t_K$ densely covering the time interval T, the sum over time instants can be replaced by the integral over T. Accordingly, the continuous generalization of the optimality criterion is of the form

$$\Delta_1(\xi) = \min_{\vartheta_2 \in \Theta_2} \int_X \mathcal{J}(\xi, \vartheta_2)\, \xi(\mathrm{d}x), \tag{6.47}$$

where X signifies a compact set where measurements are allowed, and

$$\mathcal{J}(\xi, \vartheta_2) = \int_0^{t_f} \|\eta(x,t) - \eta_2(x,t,\vartheta_2)\|^2 \mathrm{d}t. \tag{6.48}$$

The design

$$\xi^\star = \arg \max_{\xi \in \Xi(X)} \Delta_1(\xi) \tag{6.49}$$

is called a locally T-optimum design.

Thus, for this formulation, the results and algorithms outlined in Chapters 3 and 4 can be applied directly (Kuczewski *et al.*, 2004). The only additional computational effort consist in the necessity of integration with respect to time.

Example 6.3. Let consider an advection-diffusion process for a pollutant over a given area Ω, cf. Fig. 6.5. Assume that the pollutant concentration u over the time interval $T = [0, 1]$ is described by the model in the form of the partial differential equation

$$\frac{\partial u(x,t)}{\partial t} + \operatorname{div}\left(v(x)u(x,t)\right) = \operatorname{div}\left(d(x)\nabla u(x,t)\right) \quad \text{in} \quad x \in \Omega, \tag{6.50}$$

subject to initial and boundary conditions

$$\begin{cases} u(x,0) = 100\mathrm{e}^{-100(x_1^2+x_2^2)} & \text{in} \quad \Omega, \\ \dfrac{\partial u(x,t)}{\partial n} = 0 & \text{in} \quad \partial\Omega \times T, \end{cases} \tag{6.51}$$

where $\partial u/\partial n$ stands for the partial derivative of u with respect to the outward normal of boundary $\partial\Omega$. The following form of the diffusion coefficient was assumed:

$$d(x) = 0.1 + 0.1x_1^4 + 0.1x_2^4. \tag{6.52}$$

The velocity of transport medium was modelled as a radial field directed outwards with the source situated at point $P = (0, 0)$:

$$v(x) = \left(\frac{x_1^3}{\sqrt{(x_1^2 + x_2^2)}}, \frac{x_2^3}{\sqrt{(x_1^2 + x_2^2)}} \right) \tag{6.53}$$

The domain Ω, boundary $\partial\Omega$, the contour of the initial concentration of the substance $u(x, 0)$ and the velocity field of the transport medium are shown in Fig. 6.5. The alternative model has no advection part, so only the diffusion of the pollutant is considered. The model is described by the equation

$$\frac{\partial u(x,t)}{\partial t} = \operatorname{div}\left(d_a(x)\nabla u(x,t)\right) \qquad \text{in} \quad x \in \Omega \tag{6.54}$$

with boundary and initial conditions (6.51) and the same observation horizon as in the model (6.50). The spatially varying diffusion coefficient of the alternative model has the form preserving symmetry:

$$d_a(x) = \vartheta_{21} + \vartheta_{22}(x_1^4 + x_2^4). \tag{6.55}$$

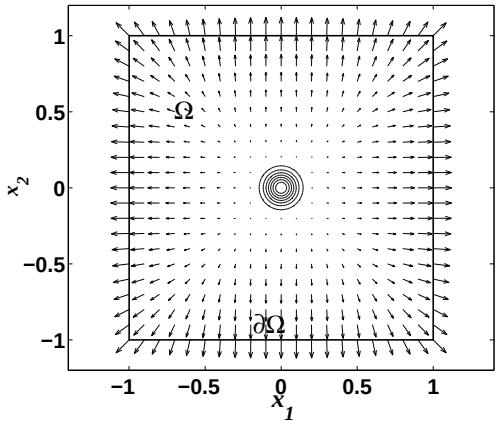


Fig. 6.5. Considered domain, its boundary, initial concentration of a pollutant (contour plot) and the velocity field of the transport medium in Example 6.3.

The allowed values of the alternative model parameters were fixed as $\vartheta_2 = (\vartheta_{21}, \vartheta_{22}) \in [0.01, 0.2]^2$. The program for computing an optimum design was written completely in the MATLAB 7.1 environment. Numerical solutions of the advection-diffusion-reaction PDE's were obtained using the finite element method (to handle that kind of PDE, a solver was written, based on the procedures from the MATLAB PDE Toolbox).

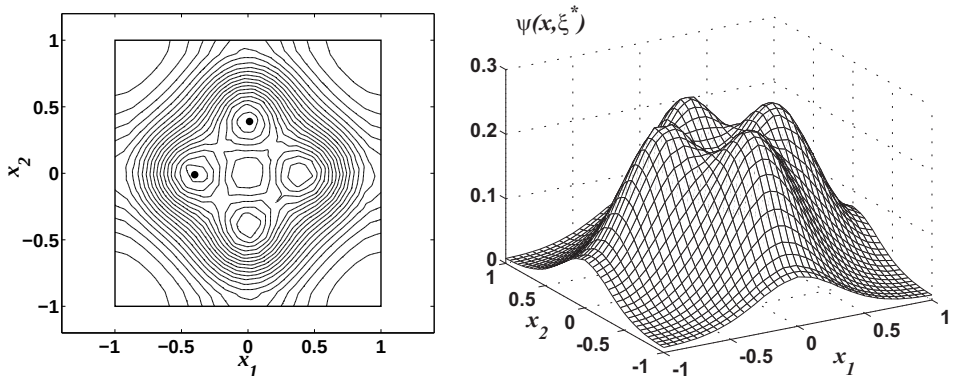
For $\epsilon = 0.002$ the Wynn-Fedorov scheme converged in 150 iterations which took about 10 hours on a PC 1.7 GHz, equipped with Pentium 4 processor and running Windows 2000. The resulting optimum design includes two points and has the form

$$\xi^\star = \begin{Bmatrix} (-0.4002, -0.0110), & (0.0120, 0.3890) \\ 0.2097, & 0.7903 \end{Bmatrix}. \tag{6.56}$$

The least profitable parameters of the alternative model are $\vartheta_2^\star = (\vartheta_{21}^\star, \vartheta_{22}^\star) = (0.1004, 0.1048)$. The obtained sensitivity function

$$\psi(x, \xi^\star) = \int_T \|\eta(x,t) - \eta_2(x,t;\vartheta_2^\star)\|^2 \mathrm{d}t \tag{6.57}$$

defining the discrepancy between the responses of both the models for the T-optimum sensor location is shown in Fig. 6.6(a). It is worth of noticing that the support points are located at the maxima of $\psi(\cdot, \xi)$ which is consistent with the presented theory. As can be seen in Fig. 6.6(b), the sensitivity function is symmetric, which results from the symmetry of the original problem. The function possesses four maxima, where the values of the T-optimum criterion equal one another, but sensors are located only at two of them. This situation is caused by unavoidable numerical inaccuracies of computations (even a very small difference between the values of the criterion is crucial for which of the points will be included



(a) (b)

Fig. 6.6. Contour plot of the sensitivity function with optimum sensor location (a) and the surface plot of the sensitivity function (b) obtained in Example 6.3.

into the current design). Note that any design concentrated at a single point being a global maximizer of the sensitivity function is T-optimal (it is clear that there are four such points, since there are four global maxima). According to Theorem 3.8, the set of T-optimal designs is convex. In these terms, any convex combination of these four single-point designs is thus T-optimal, too. As a result, the algorithm produced only one of this continuum of T-optimum designs.

6.4. Discrimination between models of DPS for correlated observations

All the problems discussed so far have been formulated in accordance with the assumption about the independence of the observations, i.e., that the random errors are not mutually correlated. This leads to elegant solutions and is very convenient because it simplifies analysis, but also has some impediments. First of all, in many practical applications, the fact that observations made at different sites are often determined by local correlations is one of the most characteristic properties of spatial data acquisition techniques. Moreover, frequently there exists no possibility of using replicated measurements, since experimental conditions are extremely difficult to reconstruct or such a procedure is unacceptable (the process of pollutant emission and proliferation in the atmosphere is a good example of such a situation). In the case of dynamic DPS's the situation is additionally complicated by the fact that random errors can be correlated in both time and space, which entails an extremely difficult nature of the problem. Thus, the majority of existing approaches concern only spatial correlations, neglecting temporal ones. Design for parameter identification involving both spatial and time correlations was considered in (Patan, 2004; Uciński, 2005). The assumption about correlations between observations leads to significant complications, since, e.g., the valuable property

of the additivity of Fisher information matrices from different measurement points is no longer justified. Thus, the information pieces from individual observations cannot be separated and direct application of the results from convex optimization is rather impossible. Moreover, replications of measurements are also not justified, since positioning any two sensors in the same location leads to the singularity of the covariance matrix. Nevertheless, it is possible to construct a numerical algorithm of the exchange type for finding locations of stationary sensors.

In the case of a design for discrimination between competing models involving mutually dependent observations, the situation is even more difficult. There are few papers concerning such a scenario for dynamic systems, and particularly for DPSs. In this context, the paper (Uciński and Bogacka, 2004) seems to be very interesting. The authors propose an approach to construct T-optimum designs for two rival multiresponse systems described by ordinary differential equations. The adopted heteroscedastic framework is based on the assumption that the observations are corrupted by a normally distributed noise with zero mean and covariance matrix which may depend on unknown parameters. The observations are not correlated in time, but correlations between different responses are admitted. Such an assumption results in a modification of the T-optimality criterion — apart from the term concerning differences between model responses, additionally a term concerning differences between covariance matrices for both the models appears. Thus, the experiment should be conducted in such a way that both the responses and the covariance matrices differ as much as possible, while the parameter values of the second model are chosen so as to make the competing model closest to the true one.

6.4.1. Heteroscedastic DPS models

The approach briefly described above can be adapted for discrimination between models of DPSs using stationary sensors. Assume that $\Omega \in \mathbb{R}^2$ stands for a bounded and simply-connected open domain with a suitably smooth boundary $\partial\Omega$. Consider the distributed parameter system whose mathematical model is described by the scalar partial differential equation

$$\frac{\partial v}{\partial t} = \mathcal{G}(x, t, v, \nabla v, \nabla^2 v), \quad (x, t) \in \Omega \times T \tag{6.58}$$

with appropriate boundary and initial conditions

$$\begin{cases} \mathcal{B}(x, t, v, \nabla v) = 0, & (x, t) \in \partial\Omega \times T, \\ \mathcal{P}(x, v, \nabla v) = 0, & (x, t) \in \Omega \times \{0\}. \end{cases} \tag{6.59}$$

where $x = (x_1, x_2) \in \Omega \cup \partial\Omega$ denotes the vector of spatial coordinates, $u(x, t) \in \mathbb{R}$ stands for the system state, $T = [0, t_f]$ is the observation horizon and $\mathcal{G}, \mathcal{B}$ and $\mathcal{P}$ are some known functions mapping their arguments into $\mathbb{R}$.

Assume that we want to optimally locate N different stationary sensors taking measurements continuously in time. If we interpret the signals acquired by the sensors as 'outputs', we can treat the considered system as a multiresponse system

with time dynamic (a system with N outputs). Then the observation equation has the form

$$y(t) = \eta(t,\xi) + \varepsilon(t), \quad t \in T, \tag{6.60}$$

where $\eta(t,\xi) = \operatorname{col}\big[v(x^1,t), v(x^2,t), \ldots, v(x^N,t)\big]$ and since we do not allow for replications, the design $\xi = \{x^1, \ldots, x^N\}$ is now interpreted as the set of sensor locations x^i, $1 = 1, \ldots, N$. The measurement noise is uncorrelated in time and $\varepsilon(t) \sim N(0, V(t,\xi))$, $\operatorname{Cov}(\varepsilon(t), \varepsilon(\tau)) = V(t,\xi)\delta(t-\tau)$, where $\delta(\,\cdot\,)$ stands for the Dirac delta function and $V(t,\xi) \in \mathbb{R}^{N\times N}$ is a symmetric and positive-definite matrix.

Two candidate models are considered:

$$\mathcal{M}_1 \,:\, E[y(t)] = \eta_1(t,\xi,\vartheta_1), \quad \operatorname{Cov}(\varepsilon(t),\varepsilon(\tau)) = V_1(t,\xi,\vartheta_1)\delta(t-\tau) \tag{6.61}$$

and

$$\mathcal{M}_2 \,:\, E[y(t)] = \eta_2(t,\xi,\vartheta_2), \quad \operatorname{Cov}(\varepsilon(t),\varepsilon(\tau)) = V_2(t,\xi,\vartheta_2)\delta(t-\tau), \tag{6.62}$$

where $\eta_\ell(t,\xi,\vartheta_\ell)$ is defined as

$$\eta_\ell(t,\xi,\vartheta_\ell) = \operatorname{col}\big[v_\ell(x^1,t), v_\ell(x^2,t), \ldots, v_\ell(x^N,t)\big], \tag{6.63}$$

$v_\ell(\,\cdot\,,\,\cdot\,)$ being the solution of the PDE

$$\frac{\partial v_\ell}{\partial t} = \mathcal{G}_\ell(x,t,v_\ell,\nabla v_\ell,\nabla^2 v_\ell,\vartheta_\ell), \quad (x,t) \in \Omega \times T \tag{6.64}$$

with boundary and initial conditions (6.59), where ϑ_ℓ stands for a vector of unknown parameters. Moreover, $V_\ell(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ have values in PD(N). Again, using T-optimality, the basic assumption we make is that model $\mathcal{M}_1$ is true, i.e., $\mathcal{G} \equiv \mathcal{G}_1$ and moreover $V(t,\xi) = V_1(t,\xi,\vartheta_1)$, $\eta(t,\xi) = \eta_1(t,\xi,\widetilde{\vartheta}_1)$ for some known value of $\widetilde{\vartheta}_1$.

Then the discrimination is performed based on the maximization of the generalized T-optimality criterion of the form

$$\Delta_1(\xi) = \min_{\vartheta_2 \in \Theta_2} \mathcal{J}(\xi,\vartheta_2), \tag{6.65}$$

where

$$\begin{aligned}\mathcal{J}(\xi,\vartheta_2) = \int_T \big\{ &\Upsilon_V(V_2^{-1}(t,\xi,\vartheta_2)) \\ &+ [\eta(t,\xi) - \eta_2(t,\xi,\vartheta_2)]^{\mathrm{T}} V_2^{-1}(t,\xi,\vartheta_2)[\eta(t,\xi) - \eta_2(t,\xi,\vartheta_2)]\big\}\,\mathrm{d}t\end{aligned} \tag{6.66}$$

and

$$\Upsilon_V(V_2^{-1}) = \operatorname{trace}(V_2^{-1}V) - \ln\det(V_2^{-1}V). \tag{6.67}$$

Such a form of the criterion results from a specific form of the logarithm of the likelihood ratio function constructed for the considered case, see (Uciński and

Bogacka, 2004) for details. Then the solution of the optimum sensor location problem reduces to finding the design

$$\xi^\star = \arg \max_{\xi \in \Xi(X)} \min_{\vartheta_2 \in \Theta_2} \mathcal{J}(\xi, \vartheta_2), \tag{6.68}$$

where $\Xi(X)$ stands now for the set of all feasible sensor configurations. In the expression (6.66) two parts can be distinguished, namely the term (6.67) dependent on the structure of the measurement noise for both the models (covariance matrices) and the term dependent on the differences between the responses of both the models. Thus we attempt to find a design for which both the parts differ as much as possible while the value of the parameter ϑ_2 is set so as to make them as close as possible.

6.4.2. Numerical approximations

Unfortunately, in the case considered we cannot provide an appropriate equivalence theorem connecting optimum solutions with properties of the sensitivity function. Thus, an adaptation of the Wynn-Fedorov numerical scheme is impossible. But we can transform the problem into the SIP one (Kuczewski, 2005), in much the same way as in Section 4.2.3. Complications arise due to the fact that evaluation of $\Delta_1(\,\cdot\,)$ involves some additional effort (an additional term appears in the criterion). On the other hand, the design now contains only support points with no corresponding weights. Generally the comments on a practical implementation given in Section 4.2.3 remain valid. The role of the efficient global optimizer should be especially pointed out, since we do not have the possibility of additionally checking the optimality of the obtained solution (such a possibility exists only when an equivalence theorem can be provided).

Additionally, it is important to impose an 'artificial' constraint preventing two different sensors being located in the same position or in a close vicinity of each other, since the covariance matrix is then singular or close to singularity.

Example 6.4. Consider the process of pollutant proliferation in the atmosphere. Its concentration v on the given area Ω is described by the following diffusion equation:

$$\frac{\partial v(x,t)}{\partial t} = \nabla^2 v(x,t) + 1000 e^{-100\|x-c\|^2} \tag{6.69}$$

with the initial and boundary conditions

$$\begin{cases} v(x,0) = 0.1, & x \in \Omega, \\ v(x,t) = 0.1(1-t), & x \in \partial\Omega \times T, \end{cases} \tag{6.70}$$

where $x \in \bar{\Omega} = [0,1]^2$, $t \in T = [0,1]$, $c = (0.25, 0.25)$. Moreover, we assume the lack of spatial correlations between the observations of pollutant concentrations, which corresponds to the covariance matrix $V = I_N$, where N stands for the number of sensors to be located. The situation described by such a model can be interpreted as the existence of a constant source of pollutant emission situated at the point c.

The alternative model is described by the equation

$$\frac{\partial v(x,t)}{\partial t} = \nabla^2 v(x,t) + 1000\vartheta_2 e^{-100\|x-d\|^2} \tag{6.71}$$

supplemented with the initial and boundary conditions (6.70), $\vartheta_2 \in \Theta_2 = [0.8, 1.2]$, $d = (0.75, 0.75)$. Moreover, the assumed form of the covariance matrix is $V_{2ij} = e^{-\rho\|x^i - x^j\|}$. Thus in the alternative model the source of pollutant emission is moved to the point with coordinates d, the intensity of emission depends on the parameter ϑ_2 and the observations of pollutant concentration are correlated. Correlation depends on the distance between the sensors and the value of the parameter ρ. In the experiment, the T-optimal positions of $N = 4$ sensors were sought. The obtained results are shown in Fig. 6.7. We can observe that sensors, initially clustered in the vicinity of the pollutant source in both the models (in the case of no correlation between the observations in the alternative model) tend to increase the mutual distances as the correlation coefficient increases.

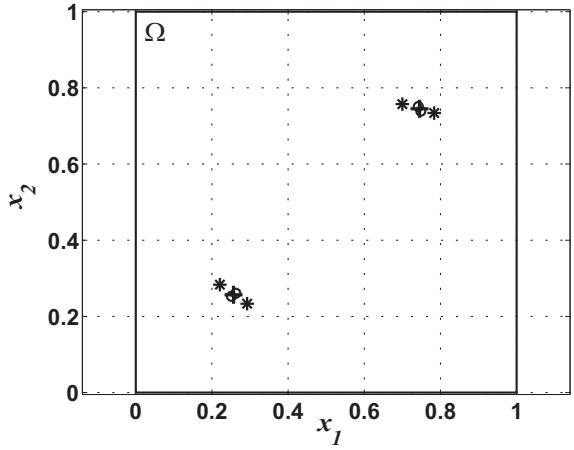


Fig. 6.7. Optimum sensor locations obtained in Example 6.4: pluses correspond to the lack of correlation in the alternative model ($V_2 = I$), open circles are positions computed for $\rho = 500$, i.e., small correlation in the alternative model, and asterisks signify the positions for $\rho = 10$, i.e., medium correlation in the alternative model.

6.5. T-optimal designs for fault detection

Methods of Fault Detection and Isolation (FDI) in dynamic systems constitute a rapidly developing part of industrial applications (Frank and Köppen-Selinger, 1997; Chen and Patton, 1999; Patton *et al.*, 2000; Chiang *et al.*, 2001; Korbicz *et al.*, 2004). However, especially in the case of processes with spatio-temporal dynamics, there are still no universal and efficient approaches to fault diagnosis.

Parameter identification is one of the basic analytical methods of fault detection when faults bring about not only changes in output signals, but also in parameter values. Usually the parameters are not measurable directly, but can be estimated based on input/output measurements. Such estimates compared with the nominal values can be used to generate residual signals. Next, analyzing such residua, a decision about an alert signalizing a fault occurrence can be made (Chen and Patton, 1999). Recently, the approach to fault detection based on adoption of optimum experimental design for parameter identification in model-based diagnosis has been presented in a series of articles (Uciński, 2003; Patan and Patan, 2003; Patan *et al.*, 2005).

The above-mentioned approach consists in an optimum location of sensors based on the so-called D_s-optimum criterion, when the parameters crucial with respect to fault detection can be separated from the remaining model parameters. It can be shown that maximization of such a criterion (accomplished by selecting appropriate sensor locations) also maximizes the power of the test constructed for verifying the null hypothesis about nominal values of system parameters ($H_0 : \vartheta = \vartheta^\star$) against the alternative one ($H_1 : \vartheta \neq \vartheta^\star$).

Another possibility of applying optimum experimental design in the field of FDI consists in applying methods of discrimination between candidate models (Baranowski and Kuczewski, 2005; Kuczewski *et al.*, 2003). The approach reduces to testing a null hypothesis corresponding to the model describing the normal state of the process functioning against the alternative one, which corresponds to an alternative model describing the faulty process. Since the maximization of the power of the test of lack of fit of the alternative model against the true one corresponds to maximization of the T-optimum criterion, it can be conjectured that a T-optimum sensor location maximizes probability of fault detection.

Example 6.5. Consider the model described by the diffusion equation in the form

$$\frac{\partial v(x,t)}{\partial t} - \nabla \cdot \nabla v(x,t) = 10\mathrm{e}^{-100\|x-c_1\|^2} + 10\mathrm{e}^{-100\|x-c_2\|^2} \tag{6.72}$$

where $x \in \Omega \cup \partial\Omega = [0,1]^2$, $t \in T = [0,1]$, $c_1 = (0.25, 0.25)$, $c_2 = (0.75, 0.75)$ with initial and boundary conditions of the form

$$\begin{cases} u(x,0) = 0.1, & x \in \Omega, \\ u(x,t) = 0.1(1-t), & (x,t) \in \partial\Omega \times T. \end{cases} \tag{6.73}$$

Such equations can be interpreted as a simple model of contamination proliferation in Ω, where v stands for the concentration of a pollutant. The driving force appearing in the model can be interpreted as two pointwise sources of pollutant emissions. They are modelled as Gaussian peaks situated at the points c_1 and c_2. The whole model (which is our reference model) simulates the normal state of the process with nominal pollutant emission by both the sources.

The alternative model differs from the reference one only by the existence of a single source of pollutant emission:

$$\frac{\partial u(x,t)}{\partial t} - \nabla \cdot \nabla v(x,t) = \vartheta_{21}\mathrm{e}^{-\vartheta_{22}\left(\|x-c_0\|^2\right)} \tag{6.74}$$

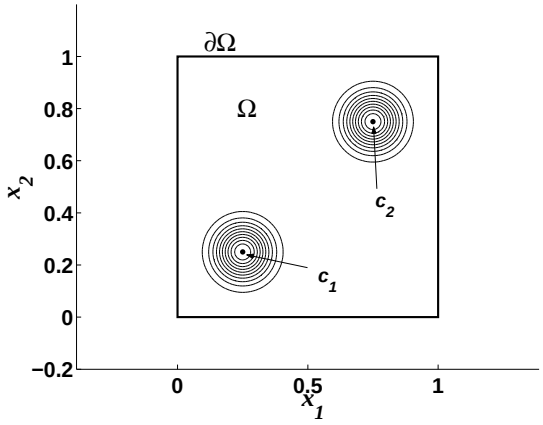


Fig. 6.8. Considered area and contour plot of the driving nominal force in the reference model of Example 6.5.

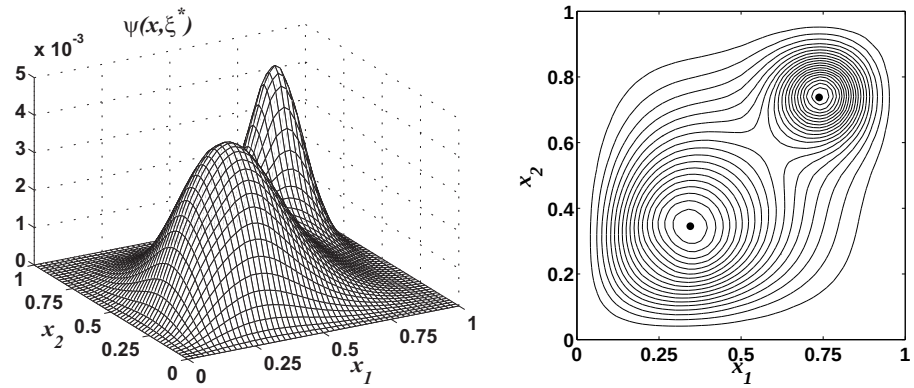


Fig. 6.9. Sensitivity function (a) and its contourplot with T-optimum sensor location (b) obtained in Example 6.5.

subject to the initial and boundary conditions (6.73), where $c_0 = (\vartheta_{23}, \vartheta_{24})$.

With an appropriate selection of the feasible range for parameters, this model can simulate the situation of a significant increase in the emission of one of the sources or a decrease in the emission of the second source. Then the problem of making a decision about the existence of such a faulty, abnormal situation reduces to discrimination between the reference and alternative models and the design of experiment can be performed as an initial step to detect and diagnose such a failure.

A Fortran 95 program to calculate the approximation of the optimum design using the Wynn-Fedorow scheme was run using the Lahey-Fujitsu Fortran

95 compiler v5.6 supplemented with IMSL library. To solve the respective global optimization problems, the ARS method was utilized. The responses of both the models were obtained using a specially implemented solver based on the Finite Element Method, and the interpolation between grid points was performed using 3D cubic splines, cf. (Uciński, 2005) for details.

It was assumed that the feasible ranges of the parameters in the alternative model were

$$\begin{aligned} 10 \le \vartheta_{21} \le 30, \qquad & 250 \le \vartheta_{22} \le 350, \\ 0.15 \le \vartheta_{23} \le 0.25, \quad & 0.15 \le \vartheta_{24} \le 0.25. \end{aligned}$$

This setting ensured that the source of pollutant emission in the alternative model was situated close to the first source in the reference model. The obtained T-optimum design has the form

$$\xi^\star = \begin{Bmatrix} (0.7377, 0.7377), & (0.3453, 0.3453) \\ 0.8889, & 0.1111 \end{Bmatrix}. \tag{6.75}$$

The plot of the resulting sensitivity function

$$\psi(x, \xi^\star) = \int_T \|\eta(x,t) - \eta_2(x,t,\vartheta_2^\star)\|^2 \,\mathrm{d}t \tag{6.76}$$

with optimum sensor locations is shown in Fig. 6.9.

The least profitable value of the alternative model parameter vector for $\xi^\star$ is

$$\vartheta_2^\star = (\vartheta_{21}^\star, \vartheta_{22}^\star, \vartheta_{23}^\star, \vartheta_{24}^\star) = (10, 250, 0.25, 0.25).$$

Thus, we can see that the sensors should be placed near both sources of pollutant emission present in the reference model, which agrees with our intuition. In those places the sensitivity function achieves its upper bound.

6.6. Summary

In this chapter we investigated possible extensions and applications of the basic formulation of the T-optimum design problem which has been studied in the previous chapters. At the beginning, a procedure was outlined to design when several competing models are possible. Clearly, the computational load increases, but the key idea of the design remains the same. Then a relatively new concept of designs with prespecified direct constraints on the design intensity (i.e., this design intensity is bounded by an atom-free measure from above) was adapted in the context of replication-free T-optimal designs. An extremely simple exchange-type algorithm was proposed to numerically determine a series of points at which single measurements are to be taken. To our knowledge, this constitutes the only treatment of what would be called the replication-free T-optimum experimental design problem.

In the remaining part of the chapter, much attention was paid to the sensor location problem for distributed parameter systems described by partial differential equations. Clearly, such models involve using sophisticated mathematical

methods, but in recompense for this effort we are in a position to describe the process more accurately and to implement more effective control strategies. Investigations in this direction had been launched by (Uciński, 2005). Here we extend his results by showing that the methodology outlined in the previous chapters is still valid in this framework and can be used to attack the problems of correlated observations (owing to the attendant serious technical difficulties, this assumption is most often avoided in the literature on optimal sensor location) and fault detection (in practice, there have been no works on this topic yet in the context of distributed parameter systems). For all the problems discussed in this chapter, the efficiency of the proposed techniques was demonstrated via several numerical examples concerning nontrivial design problems.

Chapter 7

CONCLUSIONS

From an engineering point of view it is clear that the model selection for a given system is fundamental in the sense that it determines the accuracy of the system characteristics which are identified from an identification experiment, and then the quality of the applied control scheme. On the other hand, an engineering judgement and trial-and-error analysis are quite often used to determine input and measurement ports, test signals, sampling instants, presampling filters, and some parameters such as temperature, pressure, etc., in spite of the fact that the problem has been attacked from various angles by many authors and a number of relevant results have already been reported in the literature. What is more, although it is commonly known that this area of research is difficult, since the non-linearity and nondifferentiability inherent in the experimental design for model discrimination precludes simple solution techniques, some systematic attempts at obtaining optimal experimental conditions are still made and the progress is towards more general models, more realistic assumptions and a better understanding of the nature of the measurement schedules. Logically, the number of applications should proliferate, yet this is not the case. It seems that two main reasons explain why strong formal methods are not accepted in engineering practice. First, with the use of the existing approaches, only relatively simple engineering problems can be solved without resorting to numerical methods. Second, the complexity of most existing design algorithms does not encourage engineers to apply them in practice.

Bearing this in mind, the original goal of the research reported in this dissertation was simply to develop computationally efficient methods to solve practical measurement scheduling problems for a wide class of dynamic systems described by ordinary differential and partial differential equations. In the process of executing this task, we have outlined a theoretical foundation for the adopted approach and constructed several new algorithms for various types of computation. The following is a concise summary of the contributions provided by this work to the state-of-the-art in computational techniques for optimum experimental design in dynamic systems:

- Systematizes characteristic features of the problem and analyzes the existing numerical approaches based on the classical sequential Wynn-Fedorov scheme.
- Develops an effective relaxation method for computing successive approximations to T-optimum designs (Algorithm RATO). This scheme based on the

reduction to a sequence of finite maximin problems enables us to implement the algorithms for finding optimal measurement schedules in an extremely efficient manner, without limitations to the differentiable situations as is the case for the usual Wynn-Fedorov algorithm. Simulation experiments validate the fact that making use of the proposed method may lead to dramatic gains in the values of the adopted performance index, and hence to a much greater reliability in the resulting hypothesis testing.

- Provides a smoothing method for computing approximated solutions to finite maximin problems, and characterizes the corresponding optimal measures, which allows an easy testing of any given design for optimality, and then clarifies how to adapt well-known algorithms of optimum experimental design for finding numerical approximations to the sought solutions.

- Presents the concept of replication-free designs along with a practical algorithm being a modified version of the effective method proposed by Fedorov in the context of linear regression models.

- Extends Rafajłowicz's approach of selective random search to constructing T-optimal designs and thus derives an efficient and simple computational approach based on a Markov chain Monte Carlo method.

- Incorporates an approach to constructing T-optimal designs on a finite set of allowable support points (weight-optimization algorithms).

- Formulates and solves the problem of measurement scheduling based on semi-infinite programming. Specifically, it is shown how to reduce the problem to a constrained semi-infinite problem. Then various methods can be employed to solve it numerically. It is demonstrated that the proposed approach can tackle various challenging problems of vital importance.

- Discusses the Wynn-Fedorov algorithm for the DT-optimality criterion proposed by Atkinson and provides conditions for its convergence.

- Shows how to significantly decrease the computational time indispensable to obtain numerical approximations of T-optimum designs by parallelization of computations.

- Generalizes the proposed approach to the framework of discrimination between models of distributed parameter systems, also in the presence of correlated measurement errors, and in the context of fault detection in industrial processes.

The approach suggested here has the advantage that it is independent of a particular form of the differential equation describing the dynamic system under consideration. The only requirement is the existence of sufficiently regular solutions to the state equations, and consequently highly non-linear systems can also be treated within the same framework, practically without any changes. Moreover,

it can easily be generalized to three spatial dimensions and the only limitation is the amount of required computations.

We believe that our approach has significant advantages which will make it, with sufficient development, a leading approach to solving computational problems facing engineers involved in applications of T-optimum designs.

Streszczenie

Temat rozprawy dotyczy zastosowań technik optymalnego planowania eksperymentu w identyfikacji strukturalnej systemów, ze szczególnym uwzględnieniem systemów dynamicznych, także z dynamiką czasoprzestrzenną (układy o parametrach rozłożonych).

Z inżynierskiego punktu widzenia jasne jest, że dobór odpowiedniego modelu rozważanego procesu ma kluczowe znaczenie dla wiernego oddania charakterystyki obiektu, a w konsekwencji dla jakości np. zastosowanego schematu sterowania. Z drugiej strony, ciągle jeszcze częstą praktyką jest stosowanie wiedzy eksperta lub metody prób i błędów w celu pozyskania informacji o procesie (czyli, np. doboru schematu obserwacji stanu i odpowiedzi, doboru sygnałów sterujących czy testowych itp.) ignorując fakt istnienia w literaturze wielu rezultatów proponujących systematyczne podejścia do rozwiązania tego typu problemów. Co więcej, choć powszechnie wiadomo, że optymalne planowanie eksperymentu w zadaniu dyskryminacji pomiędzy modelami jest problemem stosunkowo trudnym (pojawiające się problemy nieliniowości czy też nieróżniczkowalności kryterium powodują, że uzyskiwane rozwiązania są raczej skomplikowane) systematyczne próby w tej dziedzinie są ciągle podejmowane, a rozwój postępuje w kierunku rozpatrywania bardziej ogólnych modeli, lepiej opisujących rzeczywiste sytuacje i procesy. Powinno to zatem skutkować wzrostem liczby zastosowań aplikacyjnych. Tak jednak nie jest. Wydaje się, że przyczyny tego stanu są dwojakie. Po pierwsze, przy użyciu istniejących podejść tylko stosunkowo proste problemy inżynierskie mogą być rozpatrywane bez uciekania się do wysoce skomplikowanych metod przetwarzania i optymalizacji numerycznej. Po drugie, duża komplikacja i zawodność większości istniejących algorytmów planowania zniechęca inżynierów do stosowania ich w praktyce.

Biorąc po uwagę powyższe spostrzeżenia, głównym celem niniejszej pracy było opracowanie wydajnych, prostych w implementacji i użyciu metod rozwiązujących praktyczne problemy optymalnego doboru harmonogramu wykonywania pomiarów w zadaniu dyskryminacji pomiędzy modelami dla szerokiej klasy układów dynamicznych opisywanych równaniami różniczkowymi zwyczajnymi lub cząstkowymi. W celu zrealizowania tak postawionych założeń dokonano wyczerpującej analizy teoretycznej własności zaadoptowanego kryterium T-optymalności na bazie której skonstruowano kilka nowych algorytmów służących do numerycznego wyznaczania planów T-optymalnych w różnych sytuacjach spotykanych w praktyce.

Ponadto, zaprezentowano ciekawe podejście pozwalające na wyznaczanie tzw. planów DT-optymalnych, posiadających cenną własność łącznej maksymalizacji wiarygodności dyskryminacji pomiędzy konkurencyjnymi modelami rozpatrywanego procesu (rozumianej jako maksymalizacja mocy testu użytego w celu podjęcia decyzji o adekwatności modelu) oraz maksymalizacji jakości ocen parametrów

modelu (rozumianej jako minimalizacja wariancji uzyskanych estymatorów).

Z uwagi na poważny nakład obliczeniowy związany z koniecznością użycia metod optymalizacji globalnej do poszukiwania rozwiązań, obiecujące wydaje się zaprezentowane w pracy podejście związane ze zrównolegleniem obliczeń wykonywanych w środowisku klastra obliczeniowego.

Wyniki zawarte w pracy mają charakter ogólny i mogą być użyte bezpośrednio dla wielu klas deterministycznych, ciągłych układów dynamicznych, niezależnie od postaci opisujących je równań różniczkowych. Jedynym ograniczającym wymaganiem jest istnienie odpowiednio regularnych rozwiązań równań stanu, co pozwala na bezpośrednie zastosowanie także w przypadku wielu silnie nieliniowych systemów. Co więcej, podejście opisane w pracy może być bezpośrednio w prosty sposób uogólnione na przypadek trójwymiarowej dziedziny przestrzennej – ograniczeniem wydaje się być tutaj jedynie dostępność odpowiedniej mocy obliczeniowej.

Skuteczność zaproponowanych rozwiązań zademonstrowano na przykładzie szeregu ważnych, praktycznych problemów dyskryminacyjnych, pojawiających się np. przy opisie i modelowaniu kinetyki reakcji chemicznych czy też doborze modeli rozprzestrzeniania się zanieczyszczeń w atmosferze.

Bibliography

Antony J. (2003): *Design of Experiments for Engineers and Scientists.* — Amsterdam: Butterworth-Heinemann.

Atkinson A. C. (1992): *Optimum Experimental Designs for Parameter Estimation and for Discrimination between Models in the Presence of Prior Information.* — in V. Fedorov, W. G. Müller and I. N. Vuchkov (eds), Model-Oriented Data-Analysis: A Survey of Recent Methods. Proc. 2nd Int. IIASA Workshop in St. Kyrik, Bulgaria, May 28–June 1, 1990, Contributions to Statistics, Heidelberg: Physica-Verlag, pp. 3–30.

Atkinson A. C. (2005): *DT-Optimum Designs for Model Discrimination and Parameter Estimation.* — Research report, London School of Economics, London. Available at http://www.lse.ac.uk/collections/statistics/documents/researchreport115.pdf

Atkinson A. C. and Bailey R. A. (2001): *One Hundred Years of the Design of Experiments On and Off the Pages of* Biometrica. — Biometrika, Vol. 88, No. 1, pp. 53–97.

Atkinson A. C. and Bogacka B. (2002): *Compound and Other Optimum Designs for Systems of Nonlinear Differential Equations Arising in Chemical Kinetics.*. — Chemometrics and Intelligent Laboratory Systems, Vol. 61, pp. 17–33.

Atkinson A. C., Bogacka B. and Zhigljavsky A. A. (eds) (2001): *Optimum Design 2000.* — Dordrecht: Kluwer Scientific Publications.

Atkinson A. C. and Donev A. N. (1992): *Optimum Experimental Designs.* — Oxford: Clarendon Press.

Atkinson A. C. and Fedorov V. V. (1975a): *The Design of Experiments for Discriminating Between Two Rival Models.* — Biometrika, Vol. 62, No. 1, pp. 57–70.

Atkinson A. C. and Fedorov V. V. (1975b): *Optimal Design: Experiments for Discriminating between Several Models.* — Biometrika, Vol. 65, No. 2, pp. 289–303.

Bailey R. A. (2004): *Association Schemes: Designed Experiments, Algebra and Combinatorics.* — Cambridge: Cambridge University Press.

Banks H. T. and Kunisch K. (1989): *Estimation Techniques for Distributed Parameter Systems.* — Systems & Control: Foundations & Applications, Boston: Birkhäuser.

Banks H. T., Smith R. C. and Wang Y. (1996): *Smart Material Structures: Modeling, Estimation and Control.* — Research in Applied Mathematics, Paris: Masson.

Baranowski P. and Kuczewski B. (2005): *Fault Detection in Distributed Parameter Systems Using Parallel Processing.* — Pomiary Automatyka Kontrola, , No. 9, pp. 77–79. (in Polish)

Bard Y. (1974): *Nonlinear Parameter Estimation.* — New York: Academic Press.

Bates D. M. and Watts D. G. (1988): *Nonlinear Regression Analysis and Its Applications.* — New York: John Wiley & Sons.

Bertsekas D. P. (1999): *Nonlinear Programming.* — Optimization and Computation Series, 2nd edn, Belmont: Athena Scientific.

Bertsekas D. P. (2000): *Dynamic Programming and Optimal Control.* — Vol. 1, 2nd edn, Belmont, MA: Athena Scientific.

Biswas A. and Chaudhuri P. (2002): *An Efficient Design for Model Discrimination and Parameter Estimation.* — Biometrica, Vol. 89, No. 3, pp. 709–718.

Bogacka B. (1995): *On Information Matrices for Fixed and Random Parameters in Generally Balanced Experimental Block Designs.* — in C. P. Kitsos and W. G. Müeller (eds), MODA4 - Advances in Model Oriented Data Analysis. Proc. of the 4th International Workshop. Spetses, Greece, June 5-9, 1995, Springer Texts in Statistics, Heidelberg: Physica-Verlag.

Bookman C. (2003): *Linux Clustering: Building and Maintaining Linux Clusters.* — Indianapolis: New Riders Publishing.

Box G. E. P. and Hill W. J. (1967): *Discrimination among Mechanistics Models.* — Technometrics, Vol. 9, No. 1, pp. 57–71.

Boyd S. and Vandenberghe L. (2004): *Convex Optimization.* — Cambridge: Cambridge University Press.

Bryson A. E. (2002): *Applied Linear Optimal Control. Examples and Algorithms.* — Cambridge: Cambridge University Press.

Bryson Jr. A. E. (1999): *Dynamic Optimization.* — New York: Addison-Wesley.

Burke A. L., Duever T. A. and Pendilis A. (1994): *Model Discrimination via Designed Experiments: Discriminating between the Terminal and Penultimate Models on the Basis of Composition Data.* — Macromolecules, Vol. 27, pp. 386–399.

Burl J. B. (1999): *Linear Optimal Control. $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Methods.* — Menlo Park, CA: Addison-Wesley.

Burnham K. P. and Anderson D. R. (2002): *Model Selection and Multimodal Inference. A Practical Information-Theoretic Approach.* — 2nd edn, New York: Springer-Verlag.

Buyya R. (1999): *High Performance Cluster Computing. Architectures and Systems.* — Vol. 1Upper Saddle River: Prentice Hall.

Caines P. E. (1988): *Linear Stochastic Systems.* — New York: Wiley.

Carter M. and van Brunt B. (2000): *The Lebesgue-Stieltjes Integral. A Practical Introduction.* — New York: Springer-Verlag.

Chen J. and Patton J. R. (1999): *Robust Model-Based Fault Diagnosis for Dynamic Systems.* — Berlin: Kluwer Academic Publishers.

Chen W. H. and Seinfeld J. H. (1975): *Optimal Location of Process Measurements.* — International Journal of Control, Vol. 21, No. 6, pp. 1003–1014.

Chen X. (1998): *Global and Superlinear Convergence of Inexact Uzawa Methods for Saddle Point Problems with Nondifferentiable Mappings.* — SIAM Journal on Numerical Analysis, Vol. 35, No. 3, pp. 1130–1148.

Chen X., Qi H., Qi L. and Teo K.-l. (2004): *Smooth Convex Approximation to the Maximum Eigenvalue Function.* — Journal of Global Optimization, Vol. 30, pp. 253–270.

Chiang L. H., Russel E. L. and Braatz R. D. (2001): *Fault Detection and Diagnosis in Industrial Systems.* — London: Springer.

Cook D. and Fedorov V. V. (1995): *Constrained Optimization of Experimental Design.* — Statistics, Vol. 26, pp. 129–178.

Cox D. and Reid N. (2000): *The Theory of the Design of Experiments.* — Boca Raton: Chapman and Hall.

Dette H. and O'Brien T. E. (1999): *Optimality Criteria for Regression Models Based on Predicted Variance.* — Biometrika, Vol. 86, No. 1, pp. 93–106.

Ermakov S. M. and Zhigljavsky A. A. (1987): *Mathematical Theory of Optimal Experiments.* — Moscow: Nauka. (in Russian)

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

Fedorov V. V. (1989): *Optimal Design with Bounded Density: Optimization Algorithms of the Exchange Type.* — Journal of Statistical Planning and Inference, Vol. 22, pp. 1–13.

Fedorov V. V. and Hackl P. (1997): *Model-Oriented Design of Experiments.* — Lecture Notes in Statistics, New York: Springer-Verlag.

Fedorov V. V. and Khabarov V. (1986): *Duality of Optimal Design for Model Discrimination and Parameter Estimation.* — Biometrika, Vol. 73, No. 1, pp. 183–190.

Felsenstein K. (1992): *Optimal Bayesian Design for Discriminating among Rival Models.* — Computational Statistics & Data Analysis, Vol. 14, pp. 427–436.

Frank P. M. and Köppen-Selinger B. (1997): *New Developements Using AI in Fault Diagnosis.* — Artificial Inteligence, Vol. 10, No. 1, pp. 3–14.

Gilks W. L., Richardson S. and Spiegelharter D. J. (1995): *Markov Chain Monte Carlo in Practice.* — Boca Raton: Chapman & Hall \ CRC.

Gilmour S. G. and Trinca L. A. (2000): *Some practical advice on polynomial regression analysis from blocked response surface designs.* — Communications in Statistics - Theory and Methods, Vol. 29, pp. 2157–2180.

Givens G. H. and Hoeting J. A. (2005): *Computational Statistics.* — Wiley Series in Probability and Statistics, Hoboken New Jersey: Wiley.

Goodwin G. C. and Payne R. L. (1977): *Dynamic System Identification. Experiment Design and Data Analysis.* — Mathematics in Science and Engineering, New York: Academic Press.

Gropp W., Huss-Lederman S., Lumsdaine A., Lusk E., Nitzberg B., Saphir W. and Snir M. (1998): *MPI - The Complete Reference.* — Vol. 2, Massachusetts: MIT Press.

Gropp W., Lusk E. and Skjellum A. (1999): *Using MPI. Portable Parallel Programming with the Message Passing Interface.* — 2nd edn, Massachusetts: MIT Press.

Heiming A. and Lunze J. (1999): *Definition of Three-Tank Benchmark Problem for Controller Reconfiguration.* — Proc. 5th European Control Conference ECC'99, Karlsruhe, Germany, 31 Aug – 3 Sep, 1999, VDI/VDE-Society for Measurement and Automatic Control, 1999. Published on CD-ROM

Hettich R., Kaplan A. and Tichatschke R. (2001): *Semi-Infinite Programming: Numerical Methods.* — in C. A. Floudas and P. Pardalos (eds), Encyclopedia of Optimization, Vol. 5, Dordrecht: Kluwer Academic Publishers, pp. 112–117.

Hettich R. and Kortanek K. O. (1993): *Semi-Infinite Programming: Theory, Methods and Applications.* — SAIM Review, Vol. 35, No. 3, pp. 380–429.

Kiefer J. and Wolfowitz J. (1959): *Optimum Designs in Regression Problems.* — The Annals of Mathematical Statistics, Vol. 30, pp. 271–294.

Korbicz J., Kościelny J. M., Kowalczuk Z. and Cholewa W. (eds) (2004): *Fault Diagnosis. Models. Artificial Intelligence. Applications..* — Berlin Heidelberg: Springer-Verlag.

Kortanek K. O. and No H. (1993): *A Central Cutting Plane Algorithm for Convex Semi-Infinite Programming Ploblems.* — SIAM Journal on Optimization, Vol. 3, No. 4, pp. 901–918.

Królikowski A. (2004): *Adaptive Control with Constraints on the Control Signal.* — Poznań: Poznań University of Technology Press. (in Polish)

Królikowski A. and Eykhoff P. (1985): *Input Signals Design for System Identification: A Comparative Analysis.* — Prep. 7-th IFAC/IFORS Symp. Identification and System Parameter Estimation, York, 1985, pp. 915–920.

Kuczewski B. (2003): *Optimal Measurement Policies for Discrimination between Models of Dynamic Systems.* — in R. Kaszyński (ed.), Proc. 9-th Int. Conf. Methods and Models in Automation and Robotics, Międzyzdroje, Poland, 25–28 Aug, Szczecin: Wyd. Uczelniane Polit. Szczecińskiej, 2003, Vol. 2, pp. 1239–1244.

Kuczewski B. (2004): *Discrimination between Several Multiresponse Dynamic Models Using T-optimum Experimental Designs.* — in S. Domek and R. Kaszyński (eds), Proc. 10-th Int. Conf. Methods and Models in Automation and Robotics, Międzyzdroje, Poland, 20 Aug–2 Sep, Szczecin: Wyd. Uczelniane Polit. Szczecińskiej, 2004, Vol. 1, pp. 619–624.

Kuczewski B. (2005): *Optimum Measurement Strategies for Model Calibration of Distributed Parameter Systems.* — Pomiary, Automatyka, Kontrola, , No. 9, pp. 77–79. (in Polish)

Kuczewski B., Baranowski P. and Uciński D. (2006): *Parallel Processing in Discrimination between Models of Dynamic Systems.* — in R. Wyrzykowski, J. Dongarra, N. Meyer and J. Waśniewski (eds), Parallel Processing and Applied Mathematics, PPAM 2005, Vol. 3911 of LNCS, Berlin Heidelberg: Springer-Verlag, pp. 340–348.

Kuczewski B., Patan M. and Uciński D. (2003): *Optimal Sensor Location in Fault Detection of Distributed Parameter Systems.* — Proc. Nat. Conf. Diagnostics of Industrial Procesess, DPP, Władysławowo, Poland, 15–17 Sep 2003, Gdańsk: Pomorskie Wyd. Naukowo-Techniczne, 2003, pp. 213–218. (in Polish)

Kuczewski B., Patan M. and Uciński D. (2004): *Discrimination between Models of Distributed Parameter Systems Using T-optimum Experimental Design.* — in R. Wyrzykowski, J. Dongarra, M. Paprzycki and J. Waśniewski (eds), Parallel Processing and Applied Mathematics, PPAM 2003, Vol. 3019 of LNCS, Berlin Heidelberg: Springer-Verlag, pp. 762–769.

Kwedlo W. (2004): *A Parallel Evolutionary Algorithm for Discovery of Decision Rules.* — in R. Wyrzykowski, J. Dongarra, M. Paprzycki and J. Waśniewski (eds), Parallel Processing and Applied Mathematics, PPAM 2003, Vol. 3019 of LNCS, Berlin Heidelberg: Springer-Verlag, pp. 580–585.

Landau R. H., Fink P. J., Landau P. J. and Fink R. H. (1993): *A Scientist's and Engineer's Guide to Workstations and Supercomputers: Coping with Unix, RISC, Vectors, and Programming.* — New York: John Wiley & Sons.

MacDonald N., Minty E., Harding T. and Brown S. (2005): *Writing Message-Passing Parallel Programs with MPI. Course Notes.* — , Edinburgh Parallel Computing Centre. The University of Edinburgh. Available at http://www.epcc.ed.ac.uk/computing/training/document_archive/mpi-course/mpi-course.pdf

Mat (2005): *Optimization Toolbox for Use with MATLAB, User's Guide, Version 3.* — . Available at http://www.mathworks.com/access/helpdesk/help/toolbox/optim/optim_tb.pdf

Montgomery D. C., Peck E. A. and Vining G. G. (2001): *Introduction to Linear Regression Analysis. Third Edition.* — Wiley Series in Probability and Statistics, New York: John Wiley & Sons.

Müller W. G. (1998): *Collecting Spatial Data. Optimum Design of Experiments for Random Fields.* — Contributions to Statistics, Heidelberg: Physica-Verlag.

Müller W. G. and Ponce de Leon A. C. (1996): *Discriminating between Two Binary Data Models: Sequentially Designed Experiments.* — Journal of Statistical Computation and Simulation, Vol. 55, pp. 87–100.

O'Brien T. E. and Rawlings J. O. (1996): *A Nonsequential Design Procedure for Parameter Estimation and Model Discrimination in Nonlinear Regression Models.* — Journal of Statistical Planning and Inference, Vol. 55, pp. 77–93.

Pacheco P. S. (1997): *Programming parallel with MPI.* — San Francisco: Morgan Kaufmann.

Painter E. R. and Tits A. L. (1989): *A globally convergent algorithm with adaptively refined discretization for semi-infinite optimization in engineering design.* — IEEE Transactions on Automatic Control, Vol. 39, No. 8, pp. 903–908.

Patan M. (2004): *Optimal Observation Strategies for Parameter Estimation of Distributed Systems.* — Zielona Góra: University of Zielona Góra Press.

Patan M. and Patan K. (2003): *Calibration of Diagnostics Models for a Two Tank System Using Optimal Experimental Design.* — Proc. Nat. Conf. Diagnostics of Industrial Procesess, DPP, Władysławowo, Poland, 15–17 Sep 2003, Gdańsk: Pomorskie Wyd. Naukowo-Techniczne, 2003, pp. 189–194. (in Polish)

Patan M., Uciński D. and Baranowski P. (2005): *Optimal Observation Strategies for Fault Detection in Distributed Parameter Systems.* — Pomiary Automatyka Kontrola, , No. 9, pp. 71–73. (in Polish)

Patton R. J., Frank P. M. and Clark R. (2000): *Issues of Fault Diagnosis for Dynamic Systems.* — Berlin: Springer-Verlag.

Pázman A. (1986): *Foundations of Optimum Experimental Design.* — Mathematics and Its Applications, Dordrecht: D. Reidel Publishing Company.

Pilling M. J. and Seakins P. W. (1995): *Reaction Kinetics.* — Oxford: Oxford University Press.

Polak E. (1987): *On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design.* — SIAM Review, Vol. 29, No. 1, pp. 21–89.

Polak E. (1997): *Optimization. Algorithms and Consistent Approximations.* — Applied Mathematical Sciences, New York: Springer-Verlag.

Polak E. and He L. (1992): *Rate preserving discretization strategies for semi-infinite programming and optimal control.* — SIAM Journal on Control and Optimization, Vol. 30, No. 3, pp. 548–572.

Pollard D. (2002): *A User's Guide to Measure Theoretic Probability.* — Cambridge: Cambridge University Press.

Ponce de Leon A. C. and Atkinson A. C. (1991): *Optimum Experimental Design for Discriminating between Two Rival Models in the Presence of Prior Information.* — Biometrika, Vol. 78, No. 3, pp. 601–608.

Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. (1996): *Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing.* — 2-nd edn, Cambridge: Cambridge University Press.

Pronzato L. (2003): *Removing Non-Optimal Support Points in D-optimum Design Algorithms.* — Statistics & Probability Letters, Vol. 63, pp. 223–228.

Pukelsheim F. (1993): *Optimal Design of Experiments.* — Probability and Mathematical Statistics, New York: John Wiley & Sons.

Pukelsheim F. and Torsney B. (1991): *Optimal Weights for Experimental Design on Linearly Independent Support Points.* — Annals of Statistics, Vol. 19, No. 3, pp. 1614–1625.

Rafajłowicz E. (1978): *Design of Experiments for Parameter Identification of the Static Distributed Systems.* — Systems Science, Vol. 4, No. 4, pp. 349–361.

Rafajłowicz E. (1984): *Optimization of Actuators for Distributed Parameter Systems Identification.* — Problems of Control and Information Theory, Vol. 13, No. 1, pp. 39–51.

Rafajłowicz E. (1986): *Optimum Choice of Moving Sensor Trajectories for Distributed Parameter System Identification.* — International Journal of Control, Vol. 43, No. 5, pp. 1441–1451.

Rafajłowicz E. (1996): *Algorithms of Experimental Design with Implementations in MATHEMATICA*. — Warsaw: Akademicka Oficyna Wydawnicza PLJ. (in Polish)

Rafajłowicz E. (1998): *Selective Random Search for Optimal Experiment Designs*. — in A. C. Atkinson, L. Pronzato and H. P. Wynn (eds), MODA 5 – Advances in Model-Oriented Data Analysis and Experimental Design: Proceedings of the 5th International Workshop in Marseilles, France, 22–26 June, 1998, Contributions to Statistics, Heidelberg: Physica-Verlag, chapter 9, pp. 75–83.

Rafajłowicz E. (2006): *Optimization of Experiment with Applications in Production Quality Monitoring*. — Wrocław: Technical University Press. (in Polish)

Reemtsen R. and Rückmann J.-J. (eds) (1998): *Semi-Infinite Programming*. — Boston: Kluwer Academic Publishers.

Rockafellar R. T. (1997): *Convex Analysis*. — Princeton Landmarks in Mathematics and Physics, Princeton, New Jersey: Princeton University Press.

Schittkowski K. (2002): *Numerical Data Fitting in Dynamical Systems. A practical Introduction with Applications and Software*. — Applied Optimization, Dordrecht: Kluwer Academic Publishers.

Schwabe R. (1996): *Optimum Designs for Multi-Factor Models*. — New York: Springer-Verlag.

Seber G. A. F. and Lee A. J. (2003): *Linear Regression Analysis. Second Edition*. — Wiley Series in Probability and Statistics, Hoboken, New Jersey: John Wiley & Sons.

Seber G. A. F. and Wild C. J. (1989): *Nonlinear Regression*. — New York: John Wiley & Sons.

Shimizu K. and Aiyoshi E. (1980): *Necessary Conditions for Min-Max Problems and Algorithms by a Relaxation Procedure*. — IEEE Transactions on Automatic Control, Vol. AC-25, No. 1, pp. 62–66.

Silvey S. (1980): *Optimal design. An Introduction to the Theory for Parameter Estimation*. — London: Chapman and Hall.

Silvey S. D., Titterigton D. M. and Torsney B. (1978): *An algorithm for optimal designs on a finite design space*. — Communications in Statistics Theory and Methods A, Vol. 7, No. 14, pp. 1379–1389.

Snir M., Otto S., Huss-Lederman S., Walker D. and Dongarra J. (1998): *MPI - The Complete Reference*. — Vol. 1, Massachusetts: MIT Press.

Söderström T. and Stoica P. (1988): *System Identification*. — Series in Systems and Control Engineering, Upper Saddle River: Prentice Hall International.

Spall J. C. (2003): *Introduction to Stochastic Search and Optimization. Estimation, Simulation, and Control.* — Hoboken, NJ: John Wiley & Sons.

Stewart W. E., Shon Y. and Box G. E. P. (1998): *Discrimination and Goodness of Fit of Multiresponse Mechanistic Models.* — AIChE Journal, Vol. 44, No. 6, pp. 1404–1412.

Titterington D. M. (1976): *Algorithms for Computing D-optimum Designs on a Finite Design Space.* — Proceedings of the 1976 Conference on Information Science and Systems, Baltimore: Department of Electronic Engineering. John Hopkins University, 1976, pp. 213–216.

Tuenter H. J. H. (2001): *The Minimum L_2-Distance Projection onto the Canonical Simplex: A Simple Algorithm.* — Algo Research Quarterly, Vol. Dec 2001, pp. 53–55.

Uciński D. (1999): *Measurement Optimization for Parameter Estimation in Distributed Systems.* — Zielona Góra: Technical University Press.

Uciński D. (2003): *On optimum experimental design in technical diagnostics of industrial processes.* — Proc. Nat. Conf. Diagnostics of Industrial Procesess, DPP, Władysławowo, Poland, 15–17 Sep 2003, Gdańsk: Pomorskie Wyd. Naukowo-Techniczne, 2003, pp. 207–212. in Polish

Uciński D. (2005): *Optimal Measurement Methods for Distributed Parameter System Identification.* — Boca Raton: CRC Press.

Uciński D. and Bogacka B. (2004): *T-optimum Designs for Multiresponse Dynamic Heteroscedastic Models.* — in A. Di Bucchianico, H. Läuter and H. P. Wynn (eds), *mODa 7*, Proc. 7th Int. Workshop on Model-Oriented Data Analysis, Heeze, The Netherlands, 2004, Heidelberg: Physica-Verlag, 2004, pp. 191–199.

Uciński D. and Bogacka B. (2005): *T-optimum Designs for Discrimination between Two Multivariate Dynamic Models.* — Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 67, pp. 3–18.

van de Wal M. and de Jager B. (2001): *A Review of Methods for Input/Output Selection.* — Automatica, Vol. 37, pp. 487–510.

Žaković S. and Rustem B. (2003): *Semi-infinite Programming and Applications to Minimax Problems.* — The Annals of Opeartions Research, Vol. 124, No. 1-4, pp. 81–110.

Walter É. and Pronzato L. (1987): *Robust Experiment Design: Between Qualitative and Quantitative Identifiabilities.* — in É. Walter (ed.), Identifiability of Parametric Models, Oxford: Pergamon Press, pp. 104–113.

Walter É. and Pronzato L. (1997): *Identification of Parametric Models from Experimental Data.* — Communications and Control Engineering, Berlin: Springer-Verlag.

Waterhouse T. H., Eccleston J. A. and Duffull S. B. (2004): *On Optimal Design for Discrimination and Estimation.* — in J. Antoch (ed.), COMPSTAT 2004: Proceedings in Computational Statistics, Heidelberg: Physica-Verlag, 2004, pp. 1963–1970.

Wong W.-K. (1992): *A Unified Approach to the Construction of Minimax Designs.* — Biometrica, Vol. 79, No. 3, pp. 611–619.

Wu C.-F. and Wynn H. P. (1978): *The Convergence of General Step-Length Algorithms for Regular Optimum Design Criteria.* — The Annals of Statistics, Vol. 6, No. 6, pp. 1273–1285.

Wu S. Y., Fang S. C. and Lin C. J. (1998): *Relaxed Cutting Plane Method for Solving Linear Semi-Infinite Programming Problems.* — Journal of Optimization Theory and Applications, Vol. 99, No. 3, pp. 759–779.

Wynn H. P. (1970): *The Sequential Generation of D-optimum Experimental Designs.* — The Annals of Mathematical Statistics, Vol. 41, pp. 1655–1664.

Wyrzykowski R., Meyer N. and Stroiński M. (2004): *CLUSTERIX: National Cluster of Linux Systems.* — Proc. 2nd European Across Grids 2004 Conf., Nicosia, Cyprus, 2004. Available at: http://grid.ucy.ac.cy/axgrids04/AxGrids/index.html, http://clusterix.pcz.pl/project/publications/acrossGrids2004.pdf

Xu S. (2001): *Smoothing Method for Minimax Problems.* — Computational Optimization and Applications, Vol. 20, No. 3, pp. 267–279.

Żak S. H. (2003): *Systems and Control.* — New York: Oxford University Press.

Zarrop M. B. (1979): *Optimal Experiment Design for Dynamic System Identification.* — New York: Springer-Verlag.

Lecture Notes in Control and Computer Science

Editor-in-Chief: Józef KORBICZ

Vol. 10: Bartosz Kuczewski
Computational Aspects of Discrimination between Models of Dynamic Systems
158 p. 2006 [83-7481-030-0]

Vol. 9: Marek Kowal
Optimization of Neuro-Fuzzy Structures in Technical Diagnostics Systems
116 p. 2005 [83-89712-88-1]

Vol. 8: Wojciech Paszke
Analysis and Synthesis of Multidimensional System Classes Using Linear Matrix Inequality Methods
188 p. 2005 [83-89712-81-4]

Vol. 7: Piotr Steć
Segmentation of Colour Video Sequences Using the Fast Marching Method
110 p. 2005 [83-89712-47-4]

Vol. 6: Grzegorz Łabiak
Wykorzystanie hierarchicznego modelu współbieżnego automatu w projektowaniu sterowników cyfrowych
168 p. 2005 [83-89712-42-3]

Vol. 5: Maciej Patan
Optimal Observation Strategies for Parameter Estimation of Distributed Systems
220 p. 2004 [83-89712-03-2]

Vol. 4: Przemysław Jacewicz
Model Analysis and Synthesis of Complex Physical Systems Using Cellular Automata
134 p. 2003 [83-89321-67-X]

Vol. 3: Agnieszka Węgrzyn
Symboliczna analiza układów sterowania binarnego z wykorzystaniem wybranych metod analizy sieci Petriego
125 p. 2003 [83-89321-54-8]

Vol. 2: Grzegorz Andrzejewski
Programowy model interpretowanej sieci Petriego dla potrzeb projektowania mikrosystemów cyfrowych
109 p. 2003 [83-89321-53-X]

Vol. 1: Marcin Witczak
Identification and Fault Detection of Non-Linear Dynamic Systems
124 p. 2003 [83-88317-65-2]

Prace Naukowe z Automatyki i Informatyki

Przewodniczący: Józef KORBICZ

Tom 10: Bartosz Kuczewski
Computational Aspects of Discrimination between Models of Dynamic Systems
158 s. 2006 [83-7481-030-0]

Tom 9: Marek Kowal
Optimization of Neuro-Fuzzy Structures in Technical Diagnostics Systems
116 s. 2005 [83-89712-88-1]

Tom 8: Wojciech Paszke
Analysis and Synthesis of Multidimensional System Classes Using Linear Matrix Inequality Methods
188 s. 2005 [83-89712-81-4]

Tom 7: Piotr Steć
Segmentation of Colour Video Sequences Using the Fast Marching Method
110 s. 2005 [83-89712-47-4]

Tom 6: Grzegorz Łabiak
Wykorzystanie hierarchicznego modelu współbieżnego automatu w projektowaniu sterowników cyfrowych
168 s. 2005 [83-89712-42-3]

Tom 5: Maciej Patan
Optimal Observation Strategies for Parameter Estimation of Distributed Systems
220 s. 2004 [83-89712-03-2]

Tom 4: Przemysław Jacewicz
Model Analysis and Synthesis of Complex Physical Systems Using Cellular Automata
134 s. 2003 [83-89321-67-X]

Tom 3: Agnieszka Węgrzyn
Symboliczna analiza układów sterowania binarnego z wykorzystaniem wybranych metod analizy sieci Petriego
125 s. 2003 [83-89321-54-8]

Tom 2: Grzegorz Andrzejewski
Programowy model interpretowanej sieci Petriego dla potrzeb projektowania mikrosystemów cyfrowych
109 s. 2003 [83-89321-53-X]

Tom 1: Marcin Witczak
Identification and Fault Detection of Non-Linear Dynamic Systems
124 s. 2003 [83-88317-65-2]