# Experimental Design for Discriminating between Two Rival Models for Measurement Objectives

SOFIANE BRAHIM-BELHOUARI AND GILLES FLEURY
École Supérieure d'Électricité - Service des Mesures,
Plateau de Moulon,
91192 Gif sur Yvette, FRANCE
Tel: +33 [0]1 69 85 14 12; fax: +33 [0]1 69 85 12 34

e-mail address: Sofiane.Brahim@supelec.fr, Gilles.Fleury@supelec.fr

*Abstract:* - In this paper the problem of experimental design for a measurement purpose is studied. By extending classical approaches, the design of experiments for discriminating between two rival models is proposed. It is given by either the sequential procedure or the nonsequential Bayesian design in the presence of prior information. The approaches feasibility is finally illustrated with a simple example.

*Key-Words:* - Discrimination between models, Experimental Design, Nonlinear Estimation, Inverse Problem, Bayesian T-optimum Design.

## 1 Introduction

A careful experiment design yielding data with good information is the basis of a successful estimation procedure. The experimental design consists of searching the most informative experimental conditions for modeling objectives. Precise parameter estimation is not the only objective used for experimental design. Methods have been proposed for constructing designs for discriminating between possible model functions [1] [4] [7], and for balancing the objectives of model discrimination and precise parameter estimation.
Atkinson & Fedorov describe T-optimum design for discriminating between any number of models. Which are optimum when it is known which one of the models is true [1]. It can only be realized in practice through the use of the sequential procedure. Bayesian T-optimum design has been proposed in [9], which there is a specified prior probability that each model is true and, conditionally on this probability, prior distribution for the parameters in the two models are specified.
Techniques for discrimination should depend on the modeling purpose. This paper deals with the problem of the experimental design for discrimination in inverse problem context. We are interesting in the estimation of the measurement quantity from an observed data [5], a general formulation of a measurement is presented in section 2. Discrimination techniques for a measurement purpose are extended. Sequential experimental designs are developed for discriminating between two rival models, which need not be linear in the parameters (section 3). In the presence of prior information, nonsequential procedure is analyzed in section 4. In section 5, the both procedures are illustrated with a simple example.

## 2 Problem formulation

A problem often encountered in various domains, such as nondestructive evaluation or so-called indirect measurements, is to estimate some unknown quantity $m$ from a vector of observed values $\boldsymbol{y}$. This is due to the inability to use a transducer to measure $m$ directly or for any other reason such as harsh environment [6] [8] [11]. The general problem can be described by the following equations [5] :

$$y_i = f(x_i, \boldsymbol{\theta}) + e_i \qquad i = 1, \ldots, N, \tag{1}$$
$$m = g(\boldsymbol{\theta}). \tag{2}$$

The first one (1) is the classical nonlinear regression model, where the observation variable $\boldsymbol{y} = [y_1, \ldots, y_N]^\top$ is related to $\boldsymbol{x} = [x_1, \ldots, x_N]^\top$ the vector of the experiment design (e.g., time, frequency or space coordinates) and to the observation error $e_i$. A vector $\boldsymbol{\theta}$ ($\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$) of $p$ unknown parameters is then to be estimated from the $N$ pairs of observations $(x_i, y_i)$, where $\Theta$ is the prior feasible set for the parameters. A normal independent distribution of errors $e_i$ with zero mean and constant variance $\sigma^2$ is assumed.

The quantity to be measured, $m$, is related to the parameters $\boldsymbol{\theta}$ via (2). It is usually defined by a functional of the parametric model $m = \mathcal{G}(f)$ (i.e., involving derivation, integration, interpolation, extrapolation, ... ). This relation is then transformed into a function of $\boldsymbol{\theta}$.

The creation of candidate model structures remains mainly of a heuristic nature. So suppose that there are two rival structures of nonlinear parametric model. For the experimental conditions used, each candidate model $f_j$ $(j = 1, 2)$ generates a vector output :

$$\mathbf{f}_j(\boldsymbol{x}, \boldsymbol{\theta}_j) = [f_j(x_1, \boldsymbol{\theta}_j), \ldots, f_j(x_N, \boldsymbol{\theta}_j)]^\top, \tag{3}$$

Let $\mathbb{X}$ be the design region ($\boldsymbol{x} \in \mathbb{X}$) and $\boldsymbol{\theta}_j \in \Theta_j \subset \mathbb{R}^{p_j}$. The choice of a given structure depends on the experiment design and directly affects the estimate of $m$. This is basically related to the sensitivity of the models $f_j$ and the function $g$ to $\boldsymbol{\theta}$. This paper is concerned with the experimental design for discriminating between two rival nonlinear regression models in such a measurement framework. We consider this problem either in the absence or the presence of prior information.

## 3 Sequential experimental design

The design of experiments for discriminating between any number of models has been investigated by several authors. Atkinson and Fedorov describe T-optimum design for this purpose which are optimum when the true model is known. They assumed that one of the model is true and the experiment should then be designed to yield as large a value as possible of the sum of squares for lack of fit of the false model, this is equivalent to maximize (if we take the true model to be the first) :

$$\delta_2^{(1)}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i \{f_1(x_i, \boldsymbol{\theta}_1^*) - f_2(x_i, \hat{\boldsymbol{\theta}}_2(\boldsymbol{x}))\}^2 \tag{4}$$

we call $\delta_2^{(1)}(\boldsymbol{x})$ the noncentrality parameter of the second model, where :

$$\hat{\boldsymbol{\theta}}_2(\boldsymbol{x}) = \arg \min_{\boldsymbol{\theta}_2 \in \Theta_2} \sum_{i=1}^{n} w_i \{f_1(x_i, \boldsymbol{\theta}_1^*) - f_2(x_i, \boldsymbol{\theta}_2)\}^2$$

and $\boldsymbol{x}$ is the design :

$$\boldsymbol{x} = \{x_1, x_2, \ldots, x_n, w_1, w_2, \ldots, w_n\}$$

with $w_i = r_i/N$ specifies $n$ distinct experimental conditions and $r_i$ is the number of replicates at $x_i$. A central result is that such designs satisfay an equivalence theorem of optimum design theory, which can be used both for the construction of designs and for checking the optimality of a proposed design. The T-optimum designs described in [1] [2] can only be realized by using sequential procedure. This procedure leads to designs which are asymptotically T-optimum and which give, at each trial, the largest increase in the expected value of the sum of squares of differences between the responses from the two models. In our situation we are interested in finding the experiment design the most informative for detecting false models for a measurement purpose. In fact, each competing models $f_j (j = 1, 2)$ yields a particular evaluation of the measurement quantity $m_j$, a function of parameters ($m_j = g_j(\boldsymbol{\theta}_j)$). A direct tranposition of the experimental design using the sequential procedure [1] is to find the design $\boldsymbol{x}_m^*$ which maximizes the distance between the expected values of the measurement given by each structures :

$$\delta m(\boldsymbol{x}) = \{m_1(\hat{\boldsymbol{\theta}}_1(\boldsymbol{x})) - m_2(\hat{\boldsymbol{\theta}}_2(\boldsymbol{x}))\}^2 \tag{5}$$

In several industrials applications, the number of sensors is limited. It is due to economics requirement, harsh environment or other reasons. In this case we have to fixed the number of support points before using optimization techniques to determine the optimum design. It is strongly advise to use a global optimization techniques, since most of standard ones are unable to distinguish between a local optimum and a global one.

# 4 Bayesian design

The T-optimum design has been extended to situations in which there is a prior information [9]. The most important benefit which arises from this Bayesian approach is that optimal designs no longer depend on specific values of the parameters of the true model but only on their prior distribution, and that yields a nonsequential procedure which is very interesting in several applications.
Consider two rival models $f_1$ and $f_2$ with respective prior probabilities $\pi_1$ and $\pi_2 = 1 - \pi_1$. The set of parameters $\boldsymbol{\theta}_j$, of dimension $p_j$, has prior probability distribution $\pi_\theta(\boldsymbol{\theta}_j)(j = 1, 2)$.
The quantities

$$\delta_1^{(2)}(\boldsymbol{x}, \boldsymbol{\theta}_2) = \inf_{\theta_1 \in \Theta_1} \sum_{i=1}^{n} w_i \{f_2(x_i, \boldsymbol{\theta}_2) - f_1(x_i, \boldsymbol{\theta}_1)\}^2$$

$$\delta_2^{(1)}(\boldsymbol{x}, \boldsymbol{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^{n} w_i \{f_1(x_i, \boldsymbol{\theta}_1) - f_2(x_i, \boldsymbol{\theta}_2)\}^2$$

are the noncentrality parameters for the first model when the second is true, and vice versa. In the presence of prior information, an extension of Atkinson and Fedorov's criterion would be to find a design $\boldsymbol{x}^*$ such that

$$, (\boldsymbol{x}^*) = \sup_{x \in \mathbb{X}} , (\boldsymbol{x}) \tag{6}$$

where

$$, (\boldsymbol{x}) = \pi_1 E_{\theta_1}\{\delta_2^{(1)}(\boldsymbol{x}, \boldsymbol{\theta}_1)\} + \pi_2 E_{\theta_2}\{\delta_1^{(2)}(\boldsymbol{x}, \boldsymbol{\theta}_2)\}$$

The equivalence theorem for Bayesian T-optimum designs with prior distributions is developed in [9] to construct and check optimal designs.
A natural extension of this criterion to our problem is to find a design $\boldsymbol{x}_m^*$ which maximizes

$$, m(\boldsymbol{x}^*) = \sup_{x \in \mathbb{X}} , m(\boldsymbol{x}) \tag{7}$$

where

$$, m(\boldsymbol{x}) = \pi_1 E_{\theta_1}\{\delta m_2^{(1)}(\boldsymbol{x}, \boldsymbol{\theta}_1)\} + \pi_2 E_{\theta_2}\{\delta m_1^{(2)}(\boldsymbol{x}, \boldsymbol{\theta}_2)\}$$

with

$$\delta m_1^{(2)}(\boldsymbol{x}, \boldsymbol{\theta}_2) = \{m_2(\boldsymbol{\theta}_2) - m_1(\hat{\boldsymbol{\theta}}_1(\boldsymbol{x}))\}^2$$

$$\delta m_2^{(1)}(\boldsymbol{x}, \boldsymbol{\theta}_1) = \{m_1(\boldsymbol{\theta}_1) - m_2(\hat{\boldsymbol{\theta}}_2(\boldsymbol{x}))\}^2$$

# 5 Example

We analyze the same example for both approaches. It consists of designing an experiment over the interval $(0 \le x \le 20)$ to discriminate between the two nonlinear regression models :

$$f_1(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_1(1 - \exp(-\theta_2 \boldsymbol{x}))$$
$$f_2(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_1 \arctan(\theta_2 \boldsymbol{x})$$

the true data are supposed to come from a different structure with a normal independent distribution of errors $e_i$ with zero mean and $\sigma = 0.05$. We begin by illustrate the sequential procedure [1]. The initial design consists of trials at 0 and 20, and at each stage, the sequential design is selected by searching over a grid of 41 values of $x$ in steps of 0.5. Figure 1 shows the results of simulations of 98 trials designs in the presence of error. In the figure 2, we show the evolution of the criterion values as a function of trials number. The optimality of the proposed design was checked using the T-optimum theorem [1]. As may be seen in the figure 3, $f_1$ is the best model in the fitting criterion sense. When the number of support point is fixed, we use a global optimizer based on random search techniques [3] [10] to find out the experiment design. The support point number is assumed to be 5. After over 2000 algorithm iterations, the design for which the criterion is maximized is given by :

$$\boldsymbol{x}^* = [\ 1.5 \quad 6.2 \quad 10 \quad 19.8 \quad 20\ ]^\top$$

for which the maximum value is $5.7 \times 10^{-3}$ (figure 4).

Suppose now, we are interesting to find an experiment design to discriminate between two structures for a measurement objective, for example, the slope at the origin. It can be analytically written as $m_1 = m_2 = \theta_1\theta_2$. The search of the experiment design is done by the maximization of the criterion function $\delta m(\boldsymbol{x})$, using the global random algorithm. The design obtained for this case is :

$$\boldsymbol{x}_m^* = [\ 6 \quad 6.5 \quad 7.3 \quad 18.5 \quad 20\ ]^\top$$

and the maximized value $\delta m(\boldsymbol{x}^*) = 6.64$. The figure 5 shows the mean square fitted curves for both models and the obtained design, which is very informative for detecting false model structure for a measurement purpose.

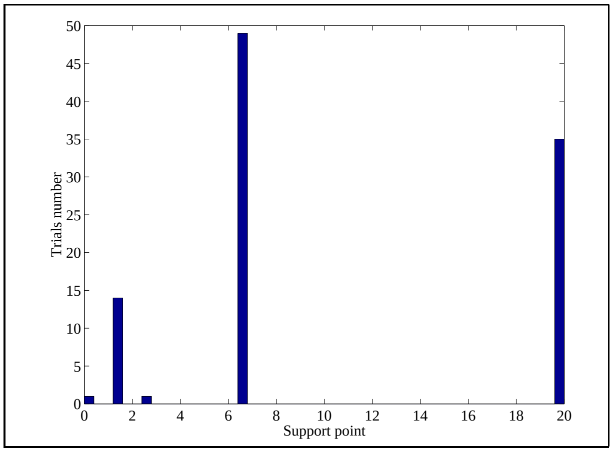


Figure 1: Obtained design.

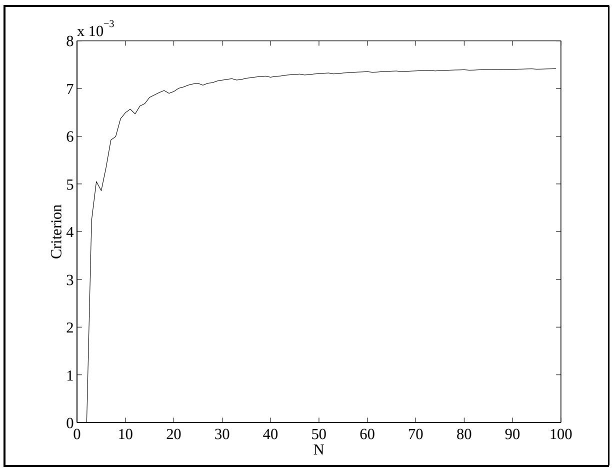


Figure 2: Evolution of criterion values as a function of trials number.

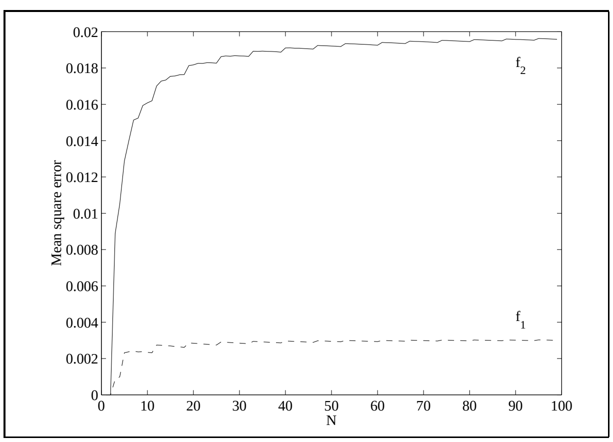


Figure 3: Evolution of the mean square error.

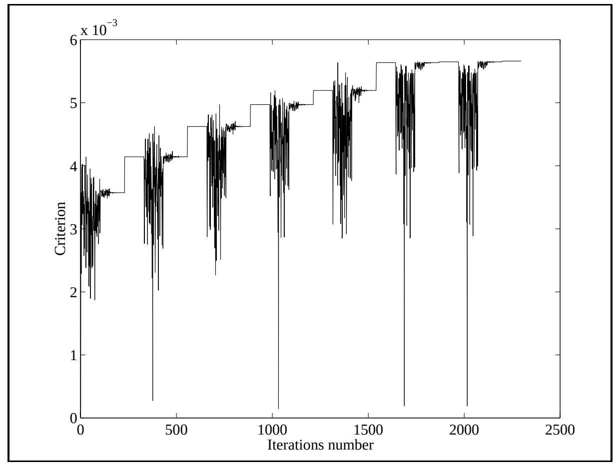


Figure 4: Evolution of criterion values as a function of iteration number.

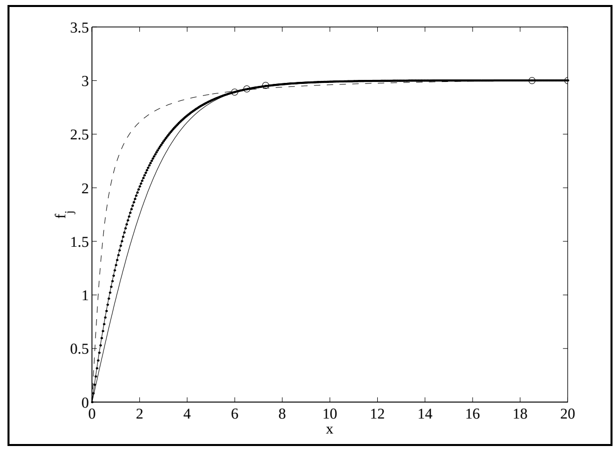


Figure 5: Data generating model (-) and mean square fitted curves for $f_1(.-.)$ and $f_2(--)$. Proposed design (o).

To study the effect of prior information on this design we consider the case where there is a prior probability distribution for the parameters of each models and a prior probability for each model to be true ($\pi_1 = \pi_2 = 0.5$). Observations from previous experiments are used to provide the prior probabilities of the parameters. The prior distribution is, in our case, normal $\pi_\theta(\boldsymbol{\theta}_j) = \mathcal{N}(\boldsymbol{\theta}_{j0}, \Omega_j)$, where $\boldsymbol{\theta}_{j0}$ is the mean and $\Omega_j$ is the covariance matrix of the parameters. The Bayesian design for a measurement objective obtained using global random algorithm with five support point is given by :

$$\boldsymbol{x}_m^* = \begin{bmatrix} 6 & 6.5 & 7.6 & 16.9 & 20 \end{bmatrix}^\top$$

and the maximized value of the criterion function , $m(\boldsymbol{x})$ is , $m(\boldsymbol{x}_m^*) = 33.9$. It might be of interest to investigate the dependence of the design and its properties on the prior distributions of the parameters.

# 6 Concluding remarks

The problem of experimental design has been considered for a measurement dedicated approach. The design of experiments for discriminating between two rivals models, in the absence and the presence of prior information, has been proposed by extending the classical methods.

The sequential experimental design for a measurement objective needs a high number of experiments, which is the main limitation of such methods. Bayesian approach yields to a nonsequential procedure, such an approach has particular disadvantage when there is great uncertainty in the prior information. A sequential design in which, observations from earlier experiments are used to update the prior probabilities of the parameters and the models, will be an alternative.

An extension of this paper results is to design for discriminating between several models for a measurement dedicated approach.

# References


[1] A.C. Atkinson and V.V. Fedorov, The design of experiments for discriminating between two rivals models, *Biometrika*, Vol.62, No.1, 1975, pp. 57-70.

[2] A.C. Atkinson and V.V. Fedorov, Optimal design : experiments for discriminating between several rivals models, *Biometrika*, Vol.62, No.2, 1975, pp. 289-303.

[3] G.A. Bekey and S.F. Masri, Random search techniques for optimization of nonlinear systems with many parameters, *Math. Comput. Simulation*, Vol.25, 1983, pp. 210-213.

[4] G.E.P. Box and W.J. Hill, Discrimination among mechanistic models, *Technometrics*, Vol.9, No.1, 1967, pp. 57-71.

[5] S. Brahim-Belhouari and G. Fleury, Probability distribustion in nonlinear estimation - a measurement dedicated approach -, *Proc. 9th IEEE Workshop on Statistical Signal and Array Processing*, Portland (USA), 1998, pp. 395-398.

[6] G. Fleury, Non-intrusive time-of-flight flow-meter-parametric estimation and optimization, *Sensors and Actuators*, Vol.A(46-47), 1995, pp. 364-368.

[7] C.Y. Huang, Planification d'expériences pour la discrimination entre structures de modèles, *Ph.D Dissertation*, Université de Paris XI, Orsay, 1991.

[8] B. D. Olin and W. Q. Meeker, Application of statistical methods to Non-Destructive Evaluation, *Technometrics*, Vol.38, No.2, 1996, pp. 95-112.

[9] A.C.M. Ponce de leon and A.C. Atkinson, Optimum experimental design for discriminating between two rivals models in the presence of prior information, *Biometrika*, Vol.78, No.3, 1991, pp. 601-608.

[10] L. Pronzato, E. Walter, A. Venot and J.-F. Lebruchec, A general purpose global optimizer : implementation and applications, *Math. Comput. Simulation*, Vol.26, 1984, pp. 412-422.

[11] L. Sandu, J. Oksman and G. Fleury, Information criteria for the choice of parametric functions for measurement, *Accepted in IEEE Trans. on Measurements*, 1999.

[12] E. Walter and L. Pronzato, Identification of parametric models from experimental data, *Springer*, London, 1997.