Model Selection via Worst-Case Criterion for Nonlinear Bounded-Error Estimation

Sofiane Brahimi-Belhouari¹, Michel Kieffer², Gilles Fleury¹,* Luc Jaulin²,³ and Éric Walter²

¹ École Supérieure d’Électricité - Service des Mesures, Plateau de Moulon, 91192 Gif-sur-Yvette, France
Phone: +33-[0]1-69-85-14-12; Fax: +33-[0]1-69-85-12-34
{Sofiane.Brahimi, Gilles.Fleury}@supelec.fr

² Laboratoire des Signaux et Systèmes, CNRS-École Supérieure d’Électricité
Plateau de Moulon, 91192 Gif-sur-Yvette, France
{kieffer, jaulin, walter}@lss.supelec.fr

³ On leave from Laboratoire d’Ingénierie des Systèmes Automatisés, Université d’Angers
2 bd Lavoisier, 49045 Angers, France
jaulin@babinet.univ-angers.fr

Abstract

In this paper the problem of model selection for measurement purpose is studied. A new selection procedure in a deterministic framework is proposed. The problem of nonlinear bounded-error estimation is viewed as a set inversion procedure. As each candidate model structure leads to a specific set of admissible values of the measurement vector, the worst-case criterion is used to select the optimal model. The selection procedure is applied to a real measurement problem, grooves dimensioning using Remote Field Eddy Current (RFEC) inspection.

1. Introduction

Data used to estimate the parameters of a model are always associated with some uncertainty. Evaluating the quality of a given estimator requires assessing quantitatively how this data uncertainty affects the estimates. It must be recognized that the results will be highly influenced by the assumptions explicitly or implicitly made about the error structure. Correct conclusions about the properties of the estimates require a correct characterization of the error structure.

The approach most commonly used is to express the errors in terms of stochastic uncertainty models. Due to incomplete information and the presence of structural model errors resulting from aggregation and obscurity of the process dynamics, a stochastic error approach is questionable, since many of these model errors are inherently deterministic [6]. In addition information about the reliability of the estimates cannot only be obtained in closed form if the error density is Gaussian and the model is linear [17] [18]. Because of these limitations, an attractive alternative to stochastic characterization has been proposed. The error in the data is no longer considered as a random variable with known or parameterized probability density function. Instead, the error is assumed to lie between some known upper and lower bounds. One is then looking for a suitable characterization of the set of all parameter vectors consistent with the model structure, data and bounds on the error.

This paper deals with the problem of model selection for measurement purpose. A general formulation of measurement in a bounded-error context is presented in Section 2. Previous works [1] [4] [16] develop some criteria leading to the choice of a nonlinear parametric model in the context of inverse problem. None of them, however, employs a deterministic framework. In this paper a new selection procedure based on bounded-error estimation is proposed. Estimation is expressed in terms of set inversion. Using the SIVIA (Set Inversion Via Interval Analysis [5]) algorithm and a recently developed image evaluation algorithm using interval computations [8], it is possible to enclose in an approximate but guaranteed way all parameter vectors that are consistent with the data and noise assumptions. The principles of these algorithms will be briefly recalled in Section 3. Thus, a set of all admissible values of the measurement vector for each candidate model structure may be computed. In
order to compare these sets, given a reference measurement, a worst-case design [7] presented in Section 4 will be used. In Section 5, the feasibility of the approach will be illustrated for a real measurement problem, grooves dimensioning using Remote Field Eddy Current (RFEC) inspection.

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 $y$. This is due to the inability to use a transducer to measure $m$ directly or for any other reason such as harsh environment [4] [14] [16]. The general problem can be described by the following equations [2]:

$$y_k = f(x_k, \theta) + e_k \quad k = 1, \ldots, N, \tag{1}$$

$$m = g(\theta). \tag{2}$$

The first one (1) is the classical nonlinear regression model, where the observation variable $y = [y_1, \ldots, y_N]^T$ is related to $x = [x_1, \ldots, x_N]^T$ the vector of the experiment design (e.g., time, frequency or space coordinates) and to the observation error $e_k$. A vector $\theta (\theta \in \Theta \subseteq \mathbb{R}^p)$ of $p$ unknown parameters is then to be estimated from the $N$ pairs of observations $(x_k, y_k)$, where $\Theta$ is the prior feasible set for the parameters.

In the context of bounded-error estimation (e.g. [12], [13], [17] and the references therein), the feasible observation error will be expressed now in terms of a set, i.e., $e_k \in \mathbb{E}_k$, where:

$$\mathbb{E}_k = \{ e_k \in \mathbb{R} : e_k \leq e_k \leq \overline{e}_k \}, \quad k = 1, \ldots, N. \tag{3}$$

In this equation $e_k$ and $\overline{e}_k$ are the lower and upper bounds on the $k$th observation error, assumed to be known. Estimating parameters in this context amounts to looking for the set $S$ of all admissible values of $\theta$ that are consistent with (1) and (3). $S$ is thus the intersection of $\Theta$ and the set of the solutions for $\theta$ of the $2N$ inequalities:

$$y_k - \overline{e}_k \leq f(x_k, \theta) \leq y_k - e_k, \quad k = 1, \ldots, N. \tag{4}$$

This bounded-error approach leads to set estimates, contrary to usual applications of the maximum-likelihood approach which yield point estimates. $S$ has been called the posterior feasible parameter set.

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

Thus, a second problem to be solved is to find the set $\mathbb{M}$ of all admissible values of $m$ corresponding to the image of $S$ using the function $g$.

The creation of candidate model structures remains mainly of a heuristic nature. So suppose that a set $F$ of $n_f$ structures of nonlinear parametric model is available. For the experimental conditions used, each candidate model $f_j (j = 1, \ldots, n_f)$ generates a vector output:

$$f_j(x, \theta) = [f_j(x_1, \theta), \ldots, f_j(x_N, \theta)]^T, \tag{5}$$

The choice of a given structure affects directly the estimate of $m$. This is basically related to the sensitivity of the models $f_j$ and function $g$ to $\theta$. This paper is concerned with the problem of model selection in such a deterministic framework.

3. Set inversion for finding parameter sets

The algorithms used for characterizing $S$ depend on whether the model output is linear in the parameters. In the first case, it is possible to use three types of algorithm involving outerbounding ellipsoids or orthotopes, or an exact description [17].

When the model is nonlinear in the parameters, the problem is much more difficult since $S$ may be nonconvex and even nonconnected. But it is still possible to get a precise description of $S$ using interval analysis and set-inversion. The task of finding all $\theta \in \Theta$ satisfying (4) may be expressed as that of characterizing the set:

$$S = \{ \theta \in \Theta : f(x, \theta) \in y - \mathbb{E} \}, \tag{6}$$

where $\mathbb{E} = \mathbb{E}_1 \times \ldots \times \mathbb{E}_N$. $S$ may be equivalently expressed as:

$$S = f^{-1}_\Theta (y - \mathbb{E}) = f^{-1}_\Theta (\mathbb{Y}), \tag{7}$$

where $f^{-1}_\Theta$ is the reciprocal function (in a set-theoretic sense) of $f$ defined over $\Theta$, and $\mathbb{Y} = y - \mathbb{E}$ is the prior feasible set for the model outputs. For a given function $f$, interval

![Figure 1. Inclusion function f_\Theta of a function f.](image-url)

analysis [11] provides an inclusion function $f_\Theta$ (see Figure 1) that returns boxes (vector intervals) that are guaranteed to contain the image by $f$ of any given box $[\theta]$ included in
the domain of $f$. The smaller the box $[\theta]$ is, the more accurately $f_\theta([\theta])$ describes $f([\theta])$ [11] [15]. This property is used in the SIVIA [5] algorithm. Suppose that $\Theta$ is a box; a box enclosing the range of $f$ over $\Theta$ is obtained evaluating $f_\theta(x, \Theta)$. If $f_\theta(x, \Theta) \cap \gamma = \emptyset$, then it is proved that no $\theta \in \Theta$ may satisfy (4) and $\Theta \cap \gamma = \emptyset$; if $f_\theta(x, \Theta) \subset \gamma$, then all values of $\theta \in \Theta$ satisfy (4), and $\Theta$ is incorporated in the approximate solution set $\gamma^*$. If none of these conditions is satisfied, $\Theta$ may be bisected into two sub-boxes on which the same tests are applied recursively. Boxes are bisected till their width is lower than some precision parameter $\epsilon$. At this step, if nothing has been proved on the box, they are stored into an uncertain set $\gamma^+$. One may thus encode $\gamma$ into two subpavings (union of nonoverlapping boxes):

$$\gamma^0 \subset \gamma \subset (\gamma^0 \cup \gamma^+)$$.

Obtaining the image $M$ of the set $\gamma$ by the function $g$ is also achieved using interval analysis by a recently developed IMAGEp (Image SubPaving evaluation) algorithm [8]. This algorithm is able to return a subpaving, i.e., a union of boxes, containing the image of a given subpaving (for example $\gamma^0$) by a function (for example $g$), with any desired precision.

4. Worst-case criterion

The theory of model building is still poor and one is usually confronted to the model selection problem, in which several mathematical functions compete to approximate empirical observations. Consider a finite number of competing models $f_j (j = 1, \ldots, n_f)$, each of which yields a particular evaluation $M_j$, of the set of all admissible values of the measurement quantity. In order to design a robust measurement estimator, we use the worst-case criterion to select the best model structure. This natural idea is used in several domains (e.g., robust control [18] and production quality [7]). We define a worst case distance:

$$\Delta m_j = \max_{m \in M_j} \| m_r - m \|$$

where $m_r$ is the reference measurement, assumed available. This information is usually obtained from experimental data or from the simulation of finite element code. The dependency of $\Delta m$ in the experimental conditions $x$ need not be made explicit since these are assumed fixed.

To choose a model according to this criterion, the model with the smallest $\Delta m$ is picked:

$$f^* = \arg \min_{1 \leq j \leq n_f} \max_{m \in M_j} \| m_r - m \|$$

5. Grooves dimensioning using Remote Field Eddy Current (RFEC) inspection

Remote field eddy current inspection is used for dimensioning grooves that may occur in ferromagnetic pipes. An efficient method has been proposed in [3] to estimate the depth and the length of corrosion grooves from measurement of a pick-up coil signal phase at different positions close to the defect. We perform a finite element calculation to obtain a parametric model of the physical phenomenon. In this approach, the dimensions of the groove are linked to the model parameters through a polynomial function. So, an estimate of the size of the groove may be computed.

Figure 2 illustrates a typical experimental apparatus used for groove dimensioning: the pipe with the sensor positioned inside the pipe and along with the coil position $x$, the phase of the detector voltage $y$ is acquired. The distance between exciter and detector coils is chosen so that the remote field condition is satisfied. We need a previous knowledge of the relation between the groove parameters (length and depth) and the observed data (detector phase). A finite element modeling has been used to obtain such a relation.

Taking into account symmetry and the range of the data, several mathematical functions may be considered to approximate it.

Three competing nonlinear model structures $f_1 (p_1 = 2)$, $f_2 (p_2 = 3)$ and $f_3 (p_3 = 4)$ are considered here:

\[
egin{align*}
f_1(x, \theta) &= \frac{\theta_1}{\theta_2 + (x - L)^2} + \frac{\theta_1}{\theta_2 + (x + L)^2} \\
 f_2(x, \theta) &= \frac{\theta_1 \arctan(\theta_2(x + \theta_3 + L))}{\theta_2 + \theta_3 + (x + L)^2} \\
 &+ \frac{\theta_1 \arctan(\theta_2(x + \theta_3 - L))}{\theta_2 - \theta_3 - (x - L)^2} \\
 f_3(x, \theta) &= \frac{\theta_1 \exp(-\theta_2|\theta_2 + \theta_4 + L|^{4/5})}{\theta_2 + \theta_3 + (x + L)^2} \\
 &+ \exp(-\theta_2|\theta_2 + \theta_4 - L|^{3/5})
\end{align*}
\]

where $L$ is an experimental constant (half-distance between coils). Figure 3 shows an example of simulated observed
data computed by a finite element code and mean-square fitted curves for each model, for given groove parameters. These model structures have a different number of parameters. The Final Prediction Error (FPE) criterion developed by Akaike (1970), measures a compromise between model complexity and goodness of fit [10]:

\[ J_{FPE}(f_j) = \frac{1 + p_j / N}{1 - p_j / N} \sum_{k=1}^{N} \left( f_j(x_k, \theta) - y_k \right)^2 \]

The obtained results are reported in Table 1. The model \( f_2 \) appears to be the best model in the FPE criterion sense (this does not still the same for another groove parameters). However, when estimating a measurement quantity, the structure classification order may change. The measurement goal is to estimate the size of the defect (length \( l \)) and depth \( d \)) from the knowledge of detector phase. In a previous approach [4] [3], the measurement quantity is expressed as a function of the optimal parameters of the model. An algebraic form for \( g \) is chosen, a bilinear polynomial function of \( \theta \), which may be written as follows:

\[ d = g_d(\theta) = \sum_{i,j} c_{ij}^d \theta_i \theta_j + \sum_i c_i^d \theta_i \]

\[ l = g_l(\theta) = \sum_{i,j} c_{ij}^l \theta_i \theta_j + \sum_i c_i^l \theta_i \]

For several real defects dimensions, the coefficients \((c_{ij}^d, c_i^d)\) and \((c_{ij}^l, c_i^l)\) have been computed in the least-squares sense [3]. Thus for each candidate models an analytical form of \( m_j = [d_j, l_j]^T \) \((j = 1, 3)\) as a function of \( \theta \) is obtained.

In deterministic framework the error is assumed to lie between some upper and lower bounds \((\bar{\varepsilon}_k, \underline{\varepsilon}_k)\). Thus each collected data from the detector belongs to an interval \( y_k \in y_k = [\bar{y}_k + \varepsilon_k, \bar{y}_k + \underline{\varepsilon}_k] \). The prior feasible set \( \mathcal{Y} \) for the model outputs is given by the cartesian product of these \( N \) intervals \( \mathcal{Y} = y_1 \times \cdots \times y_N \) (figure 4).

For our application some decisions should be taken from the numeric value of the defect size, we have to evaluate quantitatively how this data uncertainty affects the estimates. The error bounds are assumed unknown and the data uncertainty is an additional parameter to be identified. A simple bounds structure given by a constant bound \( c \) for all the data set is chosen \((y_k = [\bar{y}_k - c, \bar{y}_k + c], k = 1 \ldots N)\). Estimation procedure should find the set of the parameters values and the additional parameter \( c \). A recently developed \texttt{MEROE} (Minimal error bound estimator) algorithm [9] estimates the smallest error bound so that the set \( \mathcal{S} \) is not empty. Using the \texttt{SIVIA} algorithm the feasible set \( \mathcal{S} \) for the parameters is determined. Figure 5 shows this uncertainty information on parameters for the first model. The set \( \mathcal{M} \) of all admissible values of the measurement for each candidate models is computed using the image evaluation algorithm (Figure 6). In order to guarantee the quality of the measurement in the worst case, \( \Delta m \) for each candidate model is computed. \( f_2 \) is still the best structure in the second criterion sense. However, the selection criterion used here takes into account the final objective of modelling.

![Figure 3. Observations (o) and mean-square fitted curves for the models 1 (-), 2 (- -) and 3 (---).](image)

![Figure 4. Data with error bars.](image)

<table>
<thead>
<tr>
<th>( J_{FPE} )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2.1 \times 10^{-4}]</td>
<td>[1.9 \times 10^{-4}]</td>
<td>[0.0105]</td>
<td></td>
</tr>
<tr>
<td>( \Delta m )</td>
<td>[2.20]</td>
<td>[1.53]</td>
<td>[4.05]</td>
</tr>
</tbody>
</table>

Table 1. Obtained results for both criterion.

6. Concluding remarks

The problem of nonlinear model selection has been considered for a measurement dedicated approach. In context
of bounded-error estimation, a new selection procedure has been proposed. As each candidate structure yields a particular evaluation of the set of all admissible values of the measurement quantity, a worst case design is used to select the optimal model. The pessimism introduced by interval computations, as soon as there are many occurrences of the parameters in the formal expression of the model function, and the increasing of the complexity of the set description with the parameters number are the main limitations of such deterministic techniques. It might be of interest to use a different selection criterion and compare its results with the worst-case ones.

The choice between the stochastic approach presented in [1] [2] and the deterministic one treated in this paper, can be oriented by answering the two following questions:
- What kind of information is available?
- If the probability density function (PDF) of the additive random noise which corrupts the data is known, statistical methods are likely to be more appropriate. On the other hand, if noise specification are in terms of tolerance, the deterministic approach might be more suitable.
- What is the purpose of the measurement?
- If the goal is to guarantee the quality in the worst case, then minimax design should be preferred, which can be done in a deterministic framework such as that described here.
- If we intend to get credible PDF for the measurement, the minimization of the PDFs distance with statistical tools seems more appropriate [1].

We thank M.E. Davoust and J. Oksman for letting us use their data.

References


