Non-linear Set-membership Identification Approach based on the Bayesian Framework

Rosa M. Fernández-Cantí, Sebastian Tornil-Sin, Joaquim Blesa, and Vicenç Puig

Advanced Control Systems Research Group (SAC)
Institut de Robòtica i Informàtica Industrial, CSIC-UPC
Universitat Politècnica de Catalunya BarcelonaTech (UPC)
Llorens i Antigas 4-6, 08028 Barcelona, Spain

Monday 11th January, 2016

Abstract

This paper deals with the problem of set-membership identification of nonlinear-in-the-parameters models. To solve this problem, the paper illustrates how the Bayesian approach can be used to determine the feasible parameter set (FPS) by assuming uniform distributed estimation error and flat model prior probability distributions. The key point of the methodology is the interval evaluation of the likelihood function and the result is a set of boxes with associated credibility indexes. For each box, the credibility index is in the interval $(0, 1]$ and gives information about the amount of consistent models inside the box. The union of the boxes with credibility value equal to one provides an inner approximation of the FPS, whereas the union of all boxes provides an outer estimation. The boxes with credibility value smaller than one are located around the boundary of the FPS and their credibility index can be used to iteratively refine the inner and outer approximations up to a desired precision. The main issues and performance of the developed algorithms are discussed and illustrated by means of examples.

1 Introduction

In the Control Engineering field, the so-called Robust Identification techniques deal with the problem of obtaining not only a nominal model of the plant, but also an estimate of the uncertainty associated to the nominal model. There exist two main families of approaches to the modelling of model uncertainty:

the stochastic/probabilistic methods and the deterministic/worst case methods. For an early survey, see e.g. [1] and as precursor works see e.g. [2] and [3].

On the one hand, stochastic methods such as the classical model error model (MEM) techniques [4] and the more recent Bayesian techniques [5][6], lead to probabilistic bounds on the uncertainty region. On the other hand, deterministic methods lead to hard bounds on the uncertainty region and the most representative are the set-membership (SM) techniques (see e.g. [7][8][9]), which include the branch of interval methods (see e.g. [10][11][12][13][14][15][16][17]).

Although set-membership techniques were initially associated to deterministic approaches, in recent years many attempts have been made to blend deterministic and probabilistic estimation concepts, see for instance [18], where interval methods (deterministic) are used to approximate minimal-volume Bayesian credible (probabilistic) sets in the context of general parameter Bayesian estimation. In [19], an extension of the particle filter algorithm able to handle interval data by means of interval analysis and constraint satisfaction techniques is proposed. In [20], both deterministic and stochastic uncertainties are considered in the context of Bayesian estimation. And more recently, in the works [21] and [22], Kalman filters are designed to cope with both types of uncertainties. The research presented in this paper continues with the same spirit of blending set-membership and stochastic approaches. In particular, this paper is motivated by [23] where it is suggested a rapprochement between the set-membership and Bayesian stochastic parameter estimation approaches by establishing under which conditions both of them could be comparable.

The main contribution of this paper is to propose an algorithm that solves the set-membership parameter estimation problem for nonlinear-in-the-parameters models under the Bayesian parameter estimation framework. The key point of this algorithm is that by means of the interval evaluation of the likelihood function, and assuming a uniform distributed estimation error and flat model prior probability distributions, it produces as a result a set of boxes with associated credibility indexes. For each box, the credibility index is in the interval (0, 1] and gives information about the amount of consistent models inside the box. The union of the boxes with credibility value equal to one provides an inner approximation of the feasible parameter set (FPS), whereas the union of all boxes provides an outer estimation. The boxes with credibility value smaller than one are located around the boundary of the FPS. Compared to existing interval methods, the credibility index enhances the interval estimation process since it gives information of the amount of consistent models in each box. Consequently, it can be used to iteratively refine the inner and outer approximations of the FPS up to a desired precision. Finally, an example based on a part of a wind turbine case study is used to illustrate the proposed approach.

This paper is organized as follows: Section 2 presents the model parameterization that will be considered in this work and formulates the nonlinear set-membership parameter estimation problem. Section 3 introduces the fundamentals of the Bayesian framework. In particular, the Bayesian credible model set is defined and used to formulate, under some assumptions, the set-membership parameter estimation within the Bayesian framework. In Section 4, the parameter estimation problem is solved by means of interval evaluations. First, a basic algorithm is presented. Then, an enhanced version of the algorithm is developed in order to allow the efficient search and selection of the boxes that constitute the final FPS. In Section 5, the performance and results of the whole methodology is discussed by means of its application to a dynamic system taken from a well-known fault detection benchmark problem. Finally, Section 6 concludes the paper.

2 Problem definition

2.1 Model parametrization

Let us assume that the system can be expressed by means of the following regression model

$$y(k) = F(k, \theta) + e(k),\ k = 1, \dots, M \tag{1}$$

where:

- $F(k, \theta)$ is the regression function (or observation function), which, in a general case, is assumed to be nonlinear-in-the-parameters θ, and it can contain any function of inputs $u(k)$ and outputs $y(k)$. The regression function can be viewed as the estimation of the system response produced by the model with parameters θ, $\hat{y}(k) \equiv F(k, \theta)$.
- $\theta \in \Theta_0$ is the parameter vector of dimension $n_\theta \times 1$.
- Θ_0 is the set in the parameter space whose boundary represents the *a priori* bounds for the parameter values.
- $e(k)$ is an additive error term which is unknown but it is assumed to be bounded by a constant $|e(k)| \leq \sigma$.
- k is the discrete-time sample.

2.2 Set-membership parameter estimation problem

The set-membership parameter estimation problem consists in determining the region in the parameter space that contains all the models that are consistent with the M input/output samples [24]. This consistency parameter region is known as *Feasible Parameter Set* (FPS) and, for the model parametrization (1), it is defined as follows

$$FPS = \{\boldsymbol{\theta} \in \Theta_0 | y(k) - \sigma \leq F(k, \boldsymbol{\theta}) \leq y(k) + \sigma,\ k = 1, \ldots, M\} \tag{2}$$

In the case that the regression function is expressed linearly as $F(k, \boldsymbol{\theta}) = \boldsymbol{\varphi}^T(k)\boldsymbol{\theta}$, the exact FPS can be obtained. However, in the case that the regression function is nonlinear-in-the-parameters $\boldsymbol{\theta}$, the resulting FPS is no longer a convex polytope but a region with a much more complicated shape [24]. Thus, the exact description of the FPS may be too complex to easily deal with. Thus, several algorithms exist that obtain inner or outer simpler regions that approximate the exact FPS [24][10] leading to what are known as *Approximated Feasible Parameter Sets* (AFPS).

Inner approximations find the approximate parameter set of maximum volume such that all its parameters are inside the feasible parameter set,

$$\text{AFPS}_{\text{in}} \subseteq \text{FPS} \tag{3}$$

On the other hand, outer approximation algorithms find the approximate parameter set of minimum volume that guarantees that the feasible parameter set is inside it,

$$\text{FPS} \subseteq \text{AFPS}_{\text{out}} \tag{4}$$

When $F(k, \boldsymbol{\theta})$ is linear, boxes, parallelotopes, ellipsoids or zonotopes are used to characterize the AFPS [25][26][27]. In the nonlinear case, a minimum outer box can be determined by means of a set of optimization problems [24]. But since the parameters enter in a nonlinear way in (1), the resulting optimization problems are non-convex and obtaining the solution is NP-hard. As an alternative, the AFPS can be approximated by using interval methods such as the SIVIA (Set Inversion Via Interval Analysis) algorithm which is based on refining the initial *a priori* set Θ_0 by iteratively bisecting it [10].

3 Set-membership estimation in the Bayesian framework

3.1 Bayesian Credible Parameter Set

The parametric-type uncertainty can be described by means of the *Bayesian Credible Parameter Set* $\mathcal{B}_{\boldsymbol{\theta}}$ defined as follows:

$$\mathcal{B}_{\boldsymbol{\theta}} \equiv \{\boldsymbol{\theta} \in \mathbb{R}^{n_\theta} | p(\boldsymbol{\theta}|\mathbf{y}, \sigma) \geq c(\alpha)\} \tag{5}$$

where the process model is characterized by means of the parameter vector $\boldsymbol{\theta}$; $\mathbf{y} = (y(1), \ldots, y(M))^T$ is the measurement data vector; $c(\alpha) \in [0,1]$ is the critical value where $100(1-\alpha)\%$ is the desired credibility level; and the model posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}, \sigma)$ can be obtained by means of the Bayes' rule, $p(\boldsymbol{\theta}|\mathbf{y}, \sigma) \propto p(\mathbf{y}|\boldsymbol{\theta}, \sigma)p(\boldsymbol{\theta})$, being $p(\mathbf{y}|\boldsymbol{\theta}, \sigma)$ the likelihood of the observations $\mathbf{y}$ jointly conditioned to the model $\boldsymbol{\theta}$ and to the error bound σ.

The Bayesian credible parameter set is a very general set suitable for most identification procedures. However, for the parameter set-membership estimation problem considered here, it is not necessary to use it in its full powerfulness. A simplified version, obtained by taking the assumptions that are listed below, will be enough. Moreover, it will serve to illustrate how the deterministic set-membership parameter uncertainty region can be viewed as a particular case of the Bayesian estimation theory [23].

3.2 Assumptions

The assumptions taken in this work are the following:

First assumption: The prior distribution is flat. In the Bayesian framework, the model prior probability distribution $p(\boldsymbol{\theta})$ can be a subjective probability [28],[29]. For simplicity, here it is assumed that no information about which the value of the "true" parameter vector $\boldsymbol{\theta}$ is and consequently we take a flat $p(\boldsymbol{\theta})$ over the initial set Θ_0 . This way the model posterior distribution is directly proportional to the likelihood function of the observations, $p(\boldsymbol{\theta}|\mathbf{y}, \sigma) \propto p(\mathbf{y}|\boldsymbol{\theta}, \sigma)$, in $\boldsymbol{\Theta}_o$.

For a fixed $\boldsymbol{\theta}$, the value of $\hat{y}(k) = F(k, \boldsymbol{\theta})$, $\forall k$ can be computed. Then, for $\boldsymbol{\theta}$ and σ independent, the likelihood function $p(\mathbf{y}|\boldsymbol{\theta}, \sigma)$ coincides in form with the error term probability distribution, i.e., $p(\mathbf{y}|\boldsymbol{\theta}, \sigma) \equiv p_e(\mathbf{y} - \hat{\mathbf{y}}|\boldsymbol{\theta}, \sigma)$, where $\hat{\mathbf{y}} = (\hat{y}(1), \ldots, \hat{y}(M))^T$. The use of a non-flat prior distribution is interesting if we have reliable prior knowledge about where the FPS could could lay in $\boldsymbol{\Theta}_o$. In such a case, we could assign a prior weight to each parameter or interval in the initial set $\boldsymbol{\Theta}_o$, so the model posterior distribution would be computed as $p(\boldsymbol{\theta}|\mathbf{y}, \sigma) \propto p(\mathbf{y}|\boldsymbol{\theta}, \sigma)p(\boldsymbol{\theta})$. The additional weighting given by the prior knowledge may present advantages during the search algorithm. For instance, some points or regions could be early discarded, thus making faster the whole identification process. Moreover, the resulting probability levels in the final credible region would result in a more accurate indication of which the "true" parameter is. However, all these advantages strongly depend on the availability of good prior information. If we used wrong prior weights, the interplay with the experimental data would result in regions with biased probability levels inside. Regarding the implementation issues, the difficulty and complexity increase due to the introduction of the prior information depends on the particular algorithm. In a point-wise evaluation of the likelihood function, introducing prior distributions is trivial whereas in the interval evaluation it is more complicated.

Second assumption: The error term is uniform distributed. To obtain a hard-bounded uncertainty credible region, we must assume that the distribution of the additive error is hard-bounded. The simplest choice is to take the uniform distribution, i.e. $e(k) \sim \mathcal{U}(-\sigma, \sigma)$, where σ is selected to be the additive error bound presented in Section 2.1. In this case, the resulting likelihood function is constant and nonzero in the region where models θ are consistent with measurements and it is zero outside this region.

Note that, in this approach, we are not really concerned on obtaining the posterior distribution for a particular θ; instead, what we obtain is the region (within the initial support Θ_0) for which the posterior distribution for θ is constant and nonzero. This region corresponds to the FPS. Note also that, since the value of the posterior distribution for θ is constant over the FPS, the α value is not relevant here neither. All models θ will be equally probable to occur, and this probability value, if needed, could be obtained by forcing the volume to integrate to one in the FPS region.

Third assumption: Equation-error assumption. The likelihood function can be numerically estimated by taking the so-called *equation-error* assumption [30]. On the contrary to the *error-in-variables* approach, where the regression function itself presents an error term, the equation-error approach assumes that the error term is additive to data at each time instant k and it does not depend on k.

This way, we can assume that the error samples $e(k) = y(k) - \hat{y}(k), \forall k$, where $\hat{y}(k) = F(k, \theta)$, are i.i.d. (independent and identically distributed), and we can compute the likelihood function numerically and sample-to-sample,

$$p_e(\mathbf{y}|\theta, \sigma) = \prod_{k=1}^{M} p_e(y(k) - \hat{y}(k)|\theta, \sigma) \tag{6}$$

It is noteworthy that there is no difference in the computation of (6) whether $F(k, \theta)$ is linear in the parameters or not.

4 Computation of the Bayesian AFPS

In the Bayesian framework, it is usual to perform the computations by means of the sampling of the parameter region. In this sense Markov Chain Monte Carlo (MCMC) techniques are widely used, see e.g. [5], [6]. However, one important drawback of the point-wise characterization of the AFPS is that we cannot give guarantees that the spaces between the points belong to the AFPS. This is especially relevant in the points that are in the border since we would like to define a region enclosing the FPS in a guaranteed way. Moreover, conventional set-membership techniques allow obtaining inner and outer approximations of the FPS whereas the point-wise approaches do not enjoy this interesting feature.

In this section we present a generalization of the former point-wise methodology consisting in the interval evaluation of the likelihood function in order to obtain boxes instead of points. The new algorithms obtain not only an inner and an outer approximation of the FPS but they provide information of the amount of consistent models in each box as well.

4.1 Interval evaluation of the likelihood function

Without loose of generality, let us consider that the initial set is an axis-aligned box $[\mathbf{\Theta_0}]$ (that can be arbitrarily large if no a priori information about the FPS is available):

$$\mathbf{\Theta}_0 \triangleq [\theta_1] \times \cdots \times [\theta_{n_\theta}] \tag{7}$$

where $[\theta_i] = [\underline{\theta_i}, \overline{\theta_i}]$ are initial intervals for each parameter and the operator $\times$ represents the Cartesian product. To obtain the likelihood function over $[\mathbf{\Theta_0}]$, this box is divided into a set on N sub-boxes:

$$[\mathbf{\Theta_0}] = \left[\boldsymbol{\theta}^{\mathbf{1}}\right] \cup \cdots \cup \left[\boldsymbol{\theta}^{\mathbf{N}}\right] \tag{8}$$

where all the sub-boxes have the same size, being this size determined according to a prespecified precision vector $\varepsilon = \{\varepsilon_{\mathbf{1}}, \cdots \varepsilon_{\mathbf{n}_\theta}\}$. The total number of sub-boxes is given by the directional widths of the initial box and the values in the vector ε according to $N = \prod_{i=1}^{n_\theta} N_i$, $N_i = width([\Theta_i])/\varepsilon_i$.

Once the domain has been divided into sub-boxes, the likelihood function has to be evaluated over them. The mathematical tool that allows to evaluate functions over boxes is the interval analysis [10]. The core of interval analysis is the combination of interval arithmetic, an extension of real arithmetic to intervals, with interval extensions for elementary functions. The application of this combination allows in practice to compute the range of values that a linear or nonlinear function takes over a domain box. Unfortunately, the use of interval analysis can provide overbounded results due to the well known multi-incidence problem[1]. However, it is assured that the provided result contains all the solutions, and this is an interesting property that will be exploited by the algorithms presented in next subsections. Using interval analysis, the evaluation in a given time instant k of the regression function F over the box $\left[\boldsymbol{\theta}^{i}\right]$ results in a predicted output interval, $[\hat{y}(k)] = F(k, \left[\boldsymbol{\theta}^{i}\right])$, $[\hat{y}(k)] \triangleq [\underline{\hat{y}(k)}, \overline{\hat{y}(k)}]$.

4.2 Consistency test

Once the predicted output interval $[\hat{y}(k)]$ has been computed, it can be compared to the measurement $y(k)$ in order to obtain the interval error term $[e(k)] = y(k) - [\hat{y}(k)]$,$[e(k)] \triangleq \left[\underline{e(k)}, \overline{e(k)}\right]$. To check the consistency of $[\boldsymbol{\theta}^i]$ with the measurements, we must check the relation between the interval error $[e(k)]$

[1]Two instances of the same variable in a given interval function are treated as different variables what leads to the overestimation of the function range.

and the prespecified interval $[-\sigma, \sigma]$. This is illustrated in the Fig. 1, where the error term is assumed to be uniform-distributed, $(e(k)|\sigma) \sim U(-\sigma, \sigma)$.

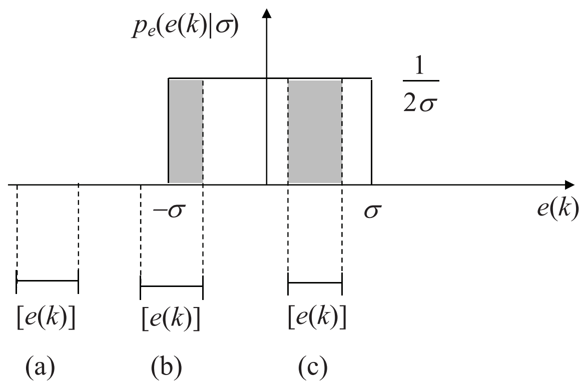


Figure 1: Consistency test

Three cases can be considered:

- If the interval error $[e(k)]$ is outside the consistency interval $[-\sigma, \sigma]$, we conclude that the interval model $[\boldsymbol{\theta}^i]$ is not consistent with the measurements (see Fig. 1(a)). Therefore these models do not belong to the FPS.
- If the interval error $[e(k)]$ is totally inside the interval $[-\sigma, \sigma]$, we conclude that the interval model $[\boldsymbol{\theta}^i]$ is consistent with the measurements and therefore it belongs to the FPS (see Fig. 1(c)).
- Finally, if the interval error $[e(k)]$ is partially inside the interval $[-\sigma, \sigma]$, we conclude that only some of the models of $[\boldsymbol{\theta}^i]$ belong to the FPS whereas others not (see Fig. 1(b)). In this later case, the interval model $[\boldsymbol{\theta}^i]$ is in the border of the FPS and it belongs to the region that is between the inner approximation of the FPS and the outer approximation.

The consistency test illustrated in Fig. 1 can be directly implemented by means of the computation of the following integration

$$I = \int_{\underline{e(k)}}^{\overline{e(k)}} p(e(k)|\sigma(k))de(k) \tag{9}$$

For the uniform case, and defining $w_{e(k)}$ as the error interval width, $w_{e(k)} \equiv \overline{e(k)} - \underline{e(k)}$, the integral (9) is $I = 0$ in the case (a), $I = w_{e(k)}/(2\sigma)$ in the case (c), and $I = \gamma w_{e(k)}/(2\sigma)$ in the case (b), with $0 < \gamma < 1$. The (a) and (c) cases can be viewed as particular cases of case (b) if we let γ take the values 0 and 1 respectively.

The value γ is called the *credibility index* and ranges from 0 to 1: $0 \leq \gamma \leq 1$. For a given box $[\theta^i]$, the value $100\gamma\%$ is related to the amount of models in the box that belong to the FPS. If $\gamma = 0$ for the box, this means that none of the models in $[\theta^i]$ belong to the FPS. If $\gamma = 1$, then the whole box $[\theta^i]$ belongs to the inner approximation of the FPS. And if γ ranges between 0 and 1, then the interval model $[\theta^i]$ belongs to the outer approximation of the FPS.

4.3 Algorithm 1

The whole procedure for set-membership identification is summarized in the Algorithm 1.

Algorithm 1 Set-membership identification using interval-based evaluation.

```
Algorithm SM-Bayes-interval-1([Θ0], ε)
  {[θ]} ← create-list([Θ0], ε)
  {γ} ← {1}
  for k = 1 to M do
    for every [θ^i] ∈ {[θ]}
      [ŷ] ← [F](k, [θ^i])
      [e] ← y(k) − [ŷ]
      if [e] ⊆ [−σ, +σ] then
        γ([θ^i]) ← γ([θ^i])
      elseif [e] ∩ [−σ, +σ] = ∅ then
        γ([θ^i]) ← 0
      else
        γ([θ^i]) ← γ([θ^i]) ∗ w ([e] ∩ [−σ, +σ]) /w ([e])
      endif
    endfor
  endfor
  normalize({γ})
  return({[θ]},{γ})
endAlgorithm
```

This algorithm uses a list of parameter boxes $\{[\boldsymbol{\theta}]\}$ and a list of credibility values $\{\gamma\}$. At the beginning of the algorithm, the list of parameter boxes is initialized with the boxes obtained by gridding the initial parameter box $[\Theta_0]$ in boxes of directional widths specified by the vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{n_\Theta})$ (according to Subsection 4.1). The list of credibility indexes $\{\gamma\}$ has the same number of elements than $\{[\boldsymbol{\theta}]\}$, each element in $\{\gamma\}$ is associated to an element in $\{[\boldsymbol{\theta}]\}$, and those elements are initialized to 1.

Two nested loops allow to deal with every sample in the data set (outer loop) and every box in $[\Theta_0]$ (inner loop). The outer loop is associated to the data samples in order to obtain a formulation of the parameter estimation algorithm that can be applied not only off-line but also on-line. In the internal code, the current box $[\boldsymbol{\theta}^i]$ is processed according to the consistency test (detailed in Subsection 4.2). After the execution of the loops, each value in $\{\gamma\}$ is in the range $[0, 1]$. A zero value indicates that none of the parameters in the associated box belong to the FPS. On the other hand, a final value equal to one indicates that this value has been obtained at every time instant, indicating that all the parameters inside the box are consistent with the whole data set. Finally, a value greater than zero and smaller than one indicates that only a certain percentage of models inside the box are consistent with the data. Hence, the result of the identification algorithm is a list of boxes with associated credibility indexes. The union of the boxes with credibility value equal to one provides an inner approximation of the FPS, whereas the union of all boxes with credibility value greater than zero provides an outer estimation. The boxes with credibility value between zero and one are located around the boundary of the FPS.

Note that in contrast to gridding methods, what we obtain here is an outer and an inner approximation for the FPS and, even more, each box inside the region between the inner and outer borders has assigned a credibility index γ.

4.4 Algorithm 2

Algorithm 1 can be improved in different ways. On the one hand, if a given parameter box $[\theta^i]$ is found to be inconsistent with the data at a given time instant k ($[e] \cap [-\sigma, +\sigma] = \emptyset$, $\gamma([\theta^i]) = 0$), then the box can be eliminated from the list to speed up the processing in the future time instants. On the other hand, it may happen that several neighbour boxes that at the end will present a credibility index equal to one could be evaluated as a whole (bigger box) obtaining the same result. In this case, the processing of just one box would be more efficient (less computing time). This claims for an alternative adaptive strategy that, instead of splitting the domain from the very beginning in as many boxes as the parameter ε indicates, it starts by considering the initial box $[\Theta_0]$ and applies a bisection process, to this initial box and to its descendants, only when is needed. Additionally, to limit the computation time, the parameter ε can be used to stop the bisection process.

Taking into account the previous discussion, the procedure summarized in Algorithm 2 is proposed. Additionally, the algorithm uses a parameter γ_{th} that will be justified later, but for a first analysis it can be assumed to be equal to one (note that, since $\gamma_k \in [0, 1]$, this is equivalent to eliminate the condition $\gamma_k \geq \gamma_{th}$ in the **elseif** statement present in the algorithm). The algorithm uses again a list of parameter boxes and a list of associated credibility indexes, but now those lists are dynamic. The processing of a given box $[\theta^i]$ at a given time instant distinguishes the three standard cases associated to the instantaneous value for the credibility index, i.e. $\gamma_k = 1$, $\gamma_k = 0$ and $0 < \gamma_k < 1$ (the function *compute-gamma* is assumed to be programmed according to the code of the **if-elseif-else** section in Algorithm 1). If $\gamma_k = 1$, the current parameter box belongs to $FPS(k)$ and its accumulated credibility is maintained. On the other hand, if $\gamma_k = 0$ then it does not belong to $FPS(k)$ and hence it also does not belong to the FPS associated to the whole set of data. Consequently, the parameter box is eliminated from the list $\{[\theta]\}$. Finally, if $0 < \gamma_k < 1$ then two sub-cases are considered. If the box is "large", it is bisected into two subboxes that are evaluated and added to the list $\{[\theta]\}$ if their computed γ_k is bigger than 0. But if the box is "small" according to the predefined ε, i.e. all the directional widths of $[\theta^i]$ are smaller than the associated values in the vector ε, then the box is not bisected and it is maintained in the list with an update of its associated accumulated γ.

Algorithm 2 Set-membership identification using interval-based evaluation.

```
Algorithm SM-Bayes-interval-2(Θ_0, ε, γ_th)
  {[θ]} ← [Θ_0]
  {γ} ← {1}
  k ← 1
  while (k ≤ M) ∧ (not isempty({[θ]})) do
    [θ^i] ← obtain-first({[θ]})
    while not isempty({[θ]}) do
      [ŷ] ← [F](k, [θ^i])
      γ_k ← compute-gamma([ŷ],y(k),σ)
      if γ_k = 1 then
        γ([θ^i]) ← γ([θ^i])
      elseif (γ_k ≥ γ_th) ∨ (width([θ^i]) ≤ ε) then
        γ([θ^i]) ← γ_k * γ([θ^i])
      else
        ([θ^l], [θ^r]) ← bisect([θ^i])
        delete-box({[θ]},[θ^i])
        [ŷ_l] ← [F](k, [θ^l])
        γ_lk ← compute-gamma([ŷ_l],y(k),σ)
        if γ_lk > 0 then
          add-box({[θ]},[θ^l])
          γ([θ^l]) ← γ_lk
        endif
        [ŷ_r] ← [F](k, [θ^r])
        γ_rk ← compute-gamma([ŷ_r],y(k),σ)
        if γ_rk > 0 then
          add-box({[θ]},[θ^r])
          γ([θ^r]) ← γ_rk
        endif
      endif
      [θ^i] ← obtain-next({[θ]})
    endwhile
  endwhile
  normalize({γ})
  return({[θ]},{γ})
endAlgorithm
```

The previously described behaviour is quite similar to the provided by the classical SIVIA algorithm [31]. However, the application of SIVIA to the parameter estimation problem presents a limitation. At a given time instant k, and in particular at the first time instant $k = 0$, SIVIA tries to obtain a good outer approximation of $FPS(k)$ according to the resolution prespecified in ε. This can imply lots of bisections for a detailed exploration of a subset of the parameter space (more precisely, an exploration of the previously computed $FPS(k-1)$) and this has associated a given computation time. However, this exploration may result unnecessary if the application of the algorithm to future data samples leads to the same final result with or without the detailed exploration in the current time instant.

This justifies the use of γ_{th} in Algorithm 2. It is used as a threshold that activates or not the bisecting process, aiming to speed up the algorithm. Hence, while processing a given box $[\theta^i]$ at a given time instant k_0, the box is not bisected if it presents a credibility index bigger or equal than the threshold, i.e. $\gamma_{k_0} > \gamma_{th}$. If in a future time instant k_1 the evaluation of box is inconsistent with the data, i.e. $\gamma_{k_1} = 0$, then the box will be excluded from the solution list. In this case, the previous decision (not bisecting $[\theta^i]$) does not alter the final result while some computations are avoided. However, on the other hand, the box can maintain an instantaneous credibility index bigger than the threshold in all time instants. In this case, the box will be accepted as belonging to the FPS even it has not an accumulated global credibility index of one. This is the price to pay, a potential overbounding of the obtained outer approximation of the FPS. However, it must be noticed that the guarantee of the obtained solution is assured. i.e. the union of the boxes returned by the algorithm constitute an outer approximation.

5 Application example

5.1 Description

In order to the illustrate the performance of the proposed approach, we are going to apply it to a well-known benchmark example from the Fault Detection and Isolation (FDI)) literature. It corresponds to a pitch system of a wind turbine [32] which dynamics can be described by a second order continuous time transfer function:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{10}$$

and the parameters to be identified are the natural frequency ω_n and the damping factor ζ. The "unknown" values for these parameters are ω_n=11.11rad/s and ζ=0.6 . Notice that the system has a nonlinear dependence with these two parameters.

In order to apply the identification methods proposed in this work, the continuous-time system (10) is discretized. According to the forward difference approximation, the transformation $s = \frac{z-1}{T_s}$ is applied to (10) to obtain the following discrete-time transfer function:

$$H(z) = \frac{\omega_n^2 T_s^2}{z^2 + (-2 + 2\zeta\omega_n T_s^2)z + (1 - 2\zeta\omega_n T_s + \omega_n^2 T_s^2)} \tag{11}$$

Then, output data $y(k)$ from the pitch system is generated by means of the simulator available with the wind turbine benchmark[2]. These data include additive sensor noise. Then, the picth system can be expressed in regressor form (1) by obtaining the difference equation from the discrete-time pitch model

[2]The benchmark is available http://www.kk-electronic.com/wind-turbine-control/competition-on-fault-detection/wind-turbine-benchmark-model.aspx

(11) as follows

$$y(k) = (2 - 2\zeta\omega_n T_s^2)y(k-1) + (-1 + 2\zeta\omega_n T_s - \omega_n^2 T_s^2)y(k-2) + (\omega_n^2 T_s^2)u(k-2) + e(k) \tag{12}$$

where the additive error term $e(k)$ takes into account the additive noise in sensors and the discretization error. The bound σ of this error can be estimated by determining the maximum (worst-case) of both.

5.2 Results

In order to perform the identification, an input signal $u(t)$ has been applied to the continuous system (10), thus generating an output signal $y(t)$ that has been corrupted by additive noise. Both signals have been sampled with a sampling time of T_s=0.0125s from t=0s to t=2s, obtaining a sequence of input/output data $\{u(k), y(k)\}_{k=1}^{M}$, with M=161. The additive error σ due to the additive noise and discretization error has been estimated and bounded by $|\sigma| \leq 0.28$. The data set used for the identification is shown in Figure 2.

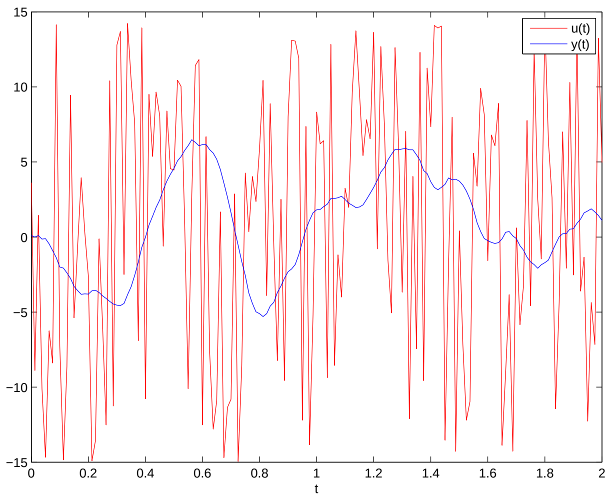
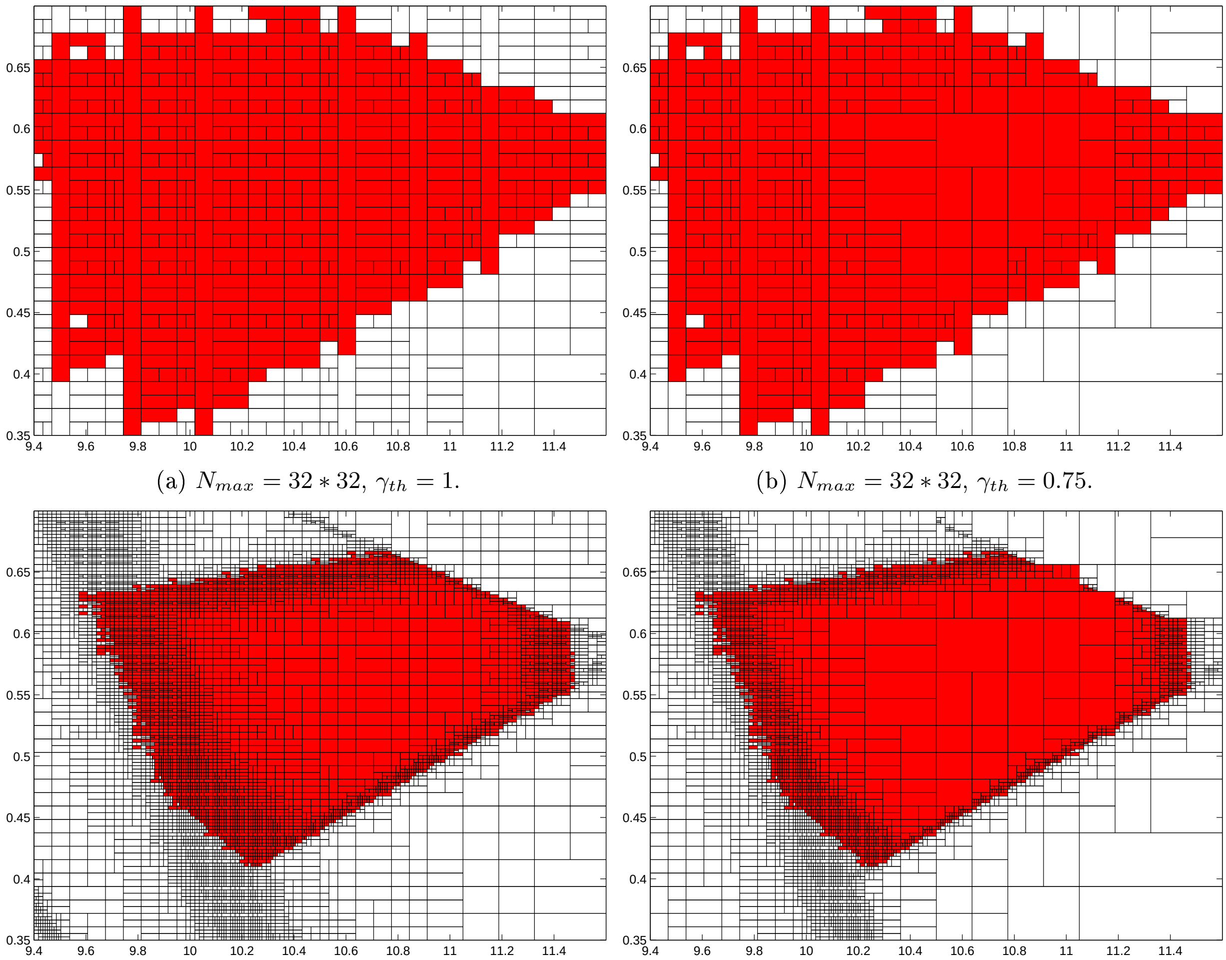


Figure 2: Input and output signals.

Algorithm 2 has been applied over the initial box $[\Theta_0] = [\underline{\omega}_n, \overline{\omega}_n] \times [\underline{\zeta}, \overline{\zeta}] = [9.4, 11.6] \times [0.35, 0.7]$, working with different resolution levels (different values of N_{max}/ε) and different values for the threshold γ_{th}. Figures 3a to 3d show some of the obtained solutions. These four displayed solution sets correspond to the use of $N_{max} = \{32 \times 32, 128 \times 128\}$ in combination with $\gamma_{th} = \{1, 0.75\}$. On one hand, it can be easily observed that the resolution of the obtained solutions mainly depends on the value for N_{max}. On the other hand, it can be observed that the use of $\gamma_{th} = 0.75$ leads to the processing of a smaller number of boxes while obtaining qualitatively similar results.

For a better analysis of the effect of the threshold γ_{th} on both the computation time and the quality



(a) $N_{max} = 32 * 32$, $\gamma_{th} = 1$.

(b) $N_{max} = 32 * 32$, $\gamma_{th} = 0.75$.

(c) $N_{max} = 128 * 128$, $\gamma_{th} = 1$.

(d) $N_{max} = 128 * 128$, $\gamma_{th} = 0.75$.

Figure 3: Some results of the application of Algorithm 2 to the application example.

of the obtained solution, more experiments have been performed and the results are summarized in Table 1. For each experiment, determined again by a given pair of values for N_{max} and γ_{th}, the computation time and the volume of the obtained solution have been determined. Regarding the computation time, it is important to indicate that Algorithm 2 has been implemented in MATLAB and executed in a laptop computer with a 2 GHz Intel Core i5 processor and 8GB of RAM memory.

Looking at Table 1, it can be verified (as it is expected) that for each value of N_{max} (any row in the table), the computation time decreases and the volume of the obtained solution increases as the value for γ_{th} decreases (going from left to right in the table). The value $\gamma_{th} = 1$ can be used as the reference value for comparison. It can be observed that the use of $\gamma_{th} = 0.7$ allows to decrease the computation time (around a 35% in average) with a really small overbounding in the obtained solutions (around 1% of increment in the volume in average). Working with $\gamma_{th} = 0.6$ allows a higher decrement of the computation time (50%) still keeping a small overbounding (still around 1%). Finally, the computation times are much more reduced (90%) by working with $\gamma_{th} = 0.5$, but with a substantial and possibly unacceptable overbounding (50%).

	$\gamma_{th} = 1$		$\gamma_{th} = 0.7$		$\gamma_{th} = 0.6$		$\gamma_{th} = 0.5$	
N_{max}	time(s)	volume	time(s)	volume	time(s)	volume	time(s)	volume
32*32	300	0.4967	239	0.4967	178	0.4997	61	0.5888
64*64	1013	0.3543	638	0.3555	501	0.3560	80	0.5207
128*128	2143	0.2978	1225	0.2991	909	0.3005	107	0.5030
256*256	3278	0.2812	1972	0.2827	1448	0.2847	130	0.4994

Table 1: Computational results of applying Algorithm 2 to the application example.

6 Conclusion

This paper has addressed the set-membership identification problem for models that are nonlinear-in-the-parameters. It has been shown how the set-membership parameter estimation problem can be reformulated such that the Bayesian framework can be used to characterize the FPS. This is possible by assuming uniform distributed error and flat model prior probability distributions.

The key point of the methodology is the computation of the likelihood function using interval methods. Hence, the likelihood function is evaluated over boxes instead of points and obtaining a measure, the so-called credibility index, of the amount of models inside each box that are consistent with the measurements. Compared to other set membership methodologies, the main contribution comes from the definition of this credibility index. Not only it gives additional information of the consistency level of each box but it also can be used to increase the efficiency in the exploration of the initial parameter space, and to refine the inner and outer approximations up to a desired precision. Moreover, the presented algorithms, basic and enhanced, serve even if the FPS is non-convex or even disjoint.

Regarding computational issues, the computation times are similar to the ones of SIVIA for a fixed precision. The main limitation is the number of parameters to deal with due to the computational complexity of this algorithm, which grows exponentially with the number of parameters. However, this drawback is common to other set membership methods. In order to overcome this drawback, the future research will focus on the introduction of more specific Bayesian tools, such as the MCMC algorithms and particle filters, to the developed algorithms that consider non-flat prior distributions and additive errors with general probability density functions.

As a further work, the effect of the interval evaluation of the parameter boxes has to be studied in detail. By directly applying interval arithmetic, and due to the multi-incidence problem, the ranges (intervals) obtained for the model output can be overbounded. On the one hand, this may increase the number of bisections. Moreover, this has an effect on the precision of the information provided by the credibility index. A possible way to explore is to combine the use of interval evaluation (through interval arithmetic) with the use of vertex evaluation, which provides an inner estimation of the range instead of an outer one. Hence, two different estimations of the model output will be obtained, leading to an interval for the credibiliy index. The use of this information in the algorithms may result of interest.

Acknowledgements

This work has been partially grant-funded by CICYT SHERECS DPI-2011-26243 and CICYT ECOCIS (Ref. DPI2013-48243-C2- 1-R) DPI-2009-13744 of the Spanish Ministry of Education and by i-Sense grant FP7-ICT-2009-6-270428 of the European Commission

References

[1] Reinelt W, Garulli A, Ljung L. Comparing different approaches to model error modelling in robust identification. Automatica. 2002;38(5):787–803.

[2] G C Goodwin BN M Gevers. Quantifying the Error in Estimated Transfer Functions with Application to Model Order Selection. IEEE Transactions on Automatic Control. 1992;.

[3] Walhberg B, Ljung L. Hard Frequency-Domain Model Error Bounds from Least-Squares Like Identification Techniques. IEEE Transactions on Automatic Control. 1992;.

[4] Ljung L. Model Validation and Model Error Modeling. In: Proceedings of the 1999 Åström Symposium on Control. Lund, Sweden; 1999. p. 15–42.

[5] Ninness B, Henriksen S. Bayesian system identification via Markov chain Monte Carlo techniques. Automatica. 2010;46(7):40–51.

[6] Schön TB, Wills A, Ninness B. System identification of nonlinear state-space models. Automatica. 2011;47(1):39–49.

[7] Milanese M, Norton JP, Piet-Lahanier H, Walter É, editors. Bounding approaches to system identification. Plennum Press; 1996.

[8] Milanese M, Taragna M. Optimality, approximation, and complexity in set membership h_∞ identification. IEEE Transactions on Automatic Control. 2002;47(10):1682–1690.

[9] Milanese M, Taragna M. H_∞ set membership identification: a survey. Automatica. 2005;41(12):2019–2032.

[10] Jaulin L, Kieffer M, Didrit O, Walter É. Applied interval analysis with examples in parameter and state estimation, robust control and robotics. Springer; 2001.

[11] Raïssi T, Ramdani N, Candau Y. Set-membership state and parameter estimation for systems described by nonlinear differential equations. Automatica. 2004;40(10):1771–1777.

[12] Milanese M, Novara C. Set-membership identification of nonlinear systems. Automatica. 2004;40(6):957–975.

[13] Raïssi T, Videau G, Zolghadri A. Interval observer design for consistency checks of nonlinear continuous-time systems. Automatica. 2010;46(3):518–527.

[14] Milanese M, Novara C. Unified Set Membership theory for identification, prediction and filtering of nonlinear systems. Automatica. 2011;47(10):2141 – 2151.

[15] Khemane F, Malti R, Raïssi T, Moreau X. Robust estimation of fractional models in the frequency domain using set membership methods. Signal Processing. 2012;(12):1591–1601.

[16] Efimov D, Perruquetti W, Raïssi T, Zolghadri A, et al. Interval Observers for Time-Varying Discrete-Time Systems. IEEE Transactions on Automatic Control. 2013;58(12):3218–3224.

[17] Efimov D, Raïssi T, Chebotarev S, Zolghadri A. Interval state observer for nonlinear time varying systems. Automatica. 2013;49(1):200–205.

[18] Jaulin L. Computing minimal-volume credible sets using interval analysis: Application to Bayesian estimation. IEEE Transactions on Signal Processing. 2006;54(9):3632–3636.

[19] Abdallah F, Gning A, Bonnifait P. Box particle filtering for nonlinear state estimation using interval analysis. Automatica. 2008;44(3):807 – 815.

[20] Klumpp V, Noack B, Baum M, Hanebeck UD. Combined set-theoretic and stochastic estimation: A comparison of the SSI and the CS filter. In: Information Fusion (FUSION), 2010 13th Conference on; 2010. p. 1–8.

[21] Noack B, Pfaff F, Hanebeck UD. Optimal Kalman gains for combined stochastic and set-membership state estimation. In: Decision and Control (CDC), 2012 IEEE 51st Annual Conference on; 2012. p. 4035–4040.

[22] Xiong J, Jauberthie C, Trave-Massuyes L, Le Gall F. Fault detection using interval Kalman filtering enhanced by constraint propagation. In: Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on; 2013. p. 490–495.

[23] Ninness BM, Goodwin GC. Rapprochement between bounded-error and stochastic estimation theory. International Journal of Adaptive Control and Signal Processing. 1995;.

[24] Milanese M, Vicino A. Estimation Theory for Nonlinear Models and Set Membership Uncertainty. In: Milanese M, Norton J, Piet-Lahanier H, Walter E, editors. Bounding Approaches to System Identification. Plenum Press; 1996. p. 345–362.

[25] Alamo T, Bravo JM, Camacho EF. Guaranteed state estimation by zonotopes. Automatica. 2005;41(6):1035–1043.

[26] Blesa J, Puig V, Saludes J. Identification for passive robust fault detection using zonotope-based set-membership approaches. International Journal of Adaptive Control and Signal Processing. 2011;25(9):788–812.

[27] Vicino A, Zappa G. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. IEEE Transaction on Automatic Control. 1996;41(6):774–785.

[28] Robert CP. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer; 2001.

[29] Jaynes ET. Probability Theory: The Logic of Science. Cambridge University Press; 2003.

[30] Sorenson HW. Least-squares estimation: from Gauss to Kalman. IEEE Spectrum. 1970;7(7):63–68.

[31] Jaulin L, Walter É. Set inversion via interval analysis for nonlinear bounded-error estimation. Automatica. 1993;29(4):1053–1064.

[32] Odgaard PF, Johnson KE. Wind turbine fault detection and fault tolerant control -a second challenge;.