



**HAL**  
open science

# Guaranteed characterization of exact confidence regions for FIR models under mild assumptions on the noise via interval analysis

Michel Kieffer, Eric Walter

## ► To cite this version:

Michel Kieffer, Eric Walter. Guaranteed characterization of exact confidence regions for FIR models under mild assumptions on the noise via interval analysis. 52nd IEEE Conference on Decision and Control CDC 2013, Dec 2013, Florence, Italy. pp.1-6, 10.1109/cdc.2013.6760681 . hal-00879946

**HAL Id: hal-00879946**

**<https://centralesupelec.hal.science/hal-00879946>**

Submitted on 5 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Guaranteed characterization of exact confidence regions for FIR models under mild assumptions on the noise via interval analysis

Michel Kieffer and Éric Walter

**Abstract**—SPS is one of the two methods proposed recently by Campi *et al.* to obtain exact, non-asymptotic confidence regions for parameter estimates under mild assumptions on the noise distribution. It does not require the measurement noise to be Gaussian (or to have any other known distribution for that matter). The numerical characterization of the resulting confidence regions is far from trivial, however, and has only been carried out so far on very low-dimensional problems via methods that could not guarantee their results and could not be extended to large-scale problems because of their intrinsic complexity. The aim of the present paper is to show how interval analysis can contribute to a guaranteed characterization of exact confidence regions in large-scale problems. The application considered is the estimation of the parameters of finite-impulse-response (FIR) models. The structure of the problem makes it possible to define a very efficient specific contractor, allowing the treatment of models with a large number of parameters, as is the rule for FIR models, and thus escaping the curse of dimensionality that often plagues interval methods.

## I. INTRODUCTION

The vector  $\mathbf{p}$  of parameters of a model is usually estimated by minimizing some cost function  $J(\mathbf{p})$ , for instance

$$J(\mathbf{p}) = \|\mathbf{y} - \mathbf{y}_m(\mathbf{p})\|_2^2, \quad (1)$$

where  $\mathbf{y}$  is a vector of data,  $\mathbf{y}_m(\mathbf{p})$  is the corresponding vector of model outputs, assumed here to be a deterministic function of  $\mathbf{p}$  and  $\|\cdot\|_2$  is a (possibly weighted)  $l_2$  norm. Then

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} J(\mathbf{p}). \quad (2)$$

Even when a single numerical vector  $\hat{\mathbf{p}}$  is obtained and  $\mathbf{y}$  and  $\mathbf{y}_m(\hat{\mathbf{p}})$  are reassuringly similar, it would be naive to consider  $\hat{\mathbf{p}}$  as the final answer to the estimation problem. One should instead attempt to attach some quality tag to  $\hat{\mathbf{p}}$  by assessing the reliability of the numerical values thus obtained. Characterizing parameter uncertainty is at the core of optimal experiment design for parameter estimation, where the most informative experimental conditions are sought.

A key issue is drawing conclusions that are as little prejudiced as possible, and this paper presents a new method for doing so in a guaranteed way, based on the SPS method recently presented by Campi and coworkers.

This work has been partly supported by the ANR CPP.

M. Kieffer and É. Walter are with the L2S, CNRS, Supelec, Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France [kieffer@lss.supelec.fr](mailto:kieffer@lss.supelec.fr), [walter@lss.supelec.fr](mailto:walter@lss.supelec.fr)

M. Kieffer is partly on leave at LTCI, CNRS, Telecom ParisTech, 46 rue Barrault, 75013 Paris and at the Institut Universitaire de France, 103 bd Saint-Michel, 75005 Paris

SPS is briefly recalled in Section II. Section III then shows how interval analysis can be used to characterize the exact confidence regions provided by this method in a global and guaranteed way. Various FIR examples are treated in Section IV, and conclusions are drawn in Section V.

## II. SPS

We assume in what follows that the system generating the data belongs to the model set, and denote its parameter vector by  $\mathbf{p}^*$ .

Exact or approximate, asymptotic or not, the characterization of parameter uncertainty via the classical approaches [6], [12], [14], [16]–[19] relies on hypotheses on the noise corrupting the data that are difficult to check from the residuals  $\mathbf{y} - \mathbf{y}_m(\hat{\mathbf{p}})$  when there are many data points, and this becomes downright impossible when there are only a few data points, as is often the case in practice. This limitation makes the new methods LSCR and SPS proposed by Campi and coworkers particularly interesting [1], [3], [4].

We shall only consider SPS here. First results on LSCR are presented in [13].

SPS [3] stands for *sign-perturbed sums*. The most striking feature of this method (which it shares with LSCR) is that it avoids a large number of the usual assumptions about the noise corrupting the data. It is not, for instance, necessary to assume that the noise is Gaussian (or that it follows any other specific probability distribution for that matter). Nor is it necessary to assume that a bound on the size of the acceptable errors is known as in bounded-error estimation. It is only assumed that the noise samples are independent, and that each of them has a probability density function that is symmetric with respect to zero. Arbitrary noise can even be dealt with if the regressors are random, independently identically and symmetrically distributed, and independent of the noise [2], [5].

SPS provides a confidence region to which  $\mathbf{p}^*$  belongs with a specified probability, by exploiting the symmetry of the noise distribution and the independence between noise samples. It is designed for linear regression, where

$$y_t = \boldsymbol{\varphi}_t^T \mathbf{p}^* + w_t, t = 1, \dots, n, \quad (3)$$

with  $\boldsymbol{\varphi}_t$  a known regression vector, which does not depend on the unknown parameters. It computes an exact confidence region for  $\mathbf{p}^*$  around the least-squares estimate  $\hat{\mathbf{p}}$ , which is

the solution to the *normal equations*

$$\sum_{t=1}^n \varphi_t (y_t - \varphi_t^T \hat{\mathbf{p}}) = \mathbf{0}. \quad (4)$$

For a generic  $\mathbf{p}$ , define

$$\mathbf{s}_0(\mathbf{p}) = \sum_{t=1}^n \varphi_t (y_t - \varphi_t^T \mathbf{p}), \quad (5)$$

and the sign-perturbed sums

$$\mathbf{s}_i(\mathbf{p}) = \sum_{t=1}^n \alpha_{i,t} \varphi_t (y_t - \varphi_t^T \mathbf{p}), \quad (6)$$

where  $i = 1, \dots, m-1$  and  $\alpha_{i,t}$  are i.i.d. random signs, i.e.,  $\alpha_{i,t} = \pm 1$  with equal probability, and

$$z_i(\mathbf{p}) = \|\mathbf{s}_i(\mathbf{p})\|_2^2, i = 0, \dots, m-1. \quad (7)$$

A confidence region  $\Sigma_q$  is obtained as the set of all values of  $\mathbf{p}$  such that  $z_0(\mathbf{p})$  is *not* among the  $q$  largest values of  $(z_i(\mathbf{p}))_{i=0}^{m-1}$ . In [3], it has been shown that  $\mathbf{p}^*$  belongs to  $\Sigma_q$  with *exact* probability  $1 - q/m$ .

$\Sigma_q$  may be defined more formally as

$$\Sigma_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m-1} \tau_i(\mathbf{p}) \geq q \right\} \quad (8)$$

where for  $i = 1, \dots, m-1$ ,

$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } z_i(\mathbf{p}) - z_0(\mathbf{p}) > 0, \\ 0 & \text{else.} \end{cases} \quad (9)$$

This is justified by the fact that if  $\sum_{i=1}^{m-1} \tau_i(\mathbf{p}) \geq q$ , then  $\tau_i(\mathbf{p}) = 1$  for at least  $q$  out of the  $m-1$  functions  $\tau_i(\mathbf{p})$ . As a consequence, there are at least  $q$  functions  $z_i(\mathbf{p})$  such that  $z_i(\mathbf{p}) > z_0(\mathbf{p})$  and  $z_0(\mathbf{p})$  is not among the  $q$  largest values of  $(z_i(\mathbf{p}))_{i=0}^{m-1}$ .

The numerical characterization of  $\Sigma_q$  is far from trivial, however, and has only been carried out so far on very low-dimensional problems via methods that could not guarantee their results and could not be extended to large-scale problems because of their intrinsic complexity. The aim of the present paper is to show how interval analysis can contribute to a guaranteed characterization of  $\Sigma_q$  in large-scale problems.

### III. GUARANTEED CHARACTERIZATION OF $\Sigma_q$ VIA INTERVAL ANALYSIS

#### A. Problem statement

Characterizing  $\Sigma_q$  may be alternatively formulated as a *set-inversion* [10] problem

$$\Sigma_q = \mathbb{P} \cap \tau^{-1}([q, m-1]), \quad (10)$$

with

$$\tau(\mathbf{p}) = \sum_{i=1}^{m-1} \tau_i(\mathbf{p}), \quad (11)$$

which may be efficiently solved via interval analysis [9], [15]. The next sections briefly recall the basic notions of interval analysis required.

#### B. Inclusion functions

Interval analysis considers closed intervals  $[x] = [\underline{x}, \bar{x}]$  of  $\mathbb{R}$  and extends all arithmetic operations and elementary functions on real numbers to these intervals.

For arithmetic operations,

$$[x] \circ [y] = \{x \circ y \mid x \in [x], y \in [y]\}, \quad (12)$$

where  $\circ \in \{+, -, \cdot, /\}$ . The set (12) is easily evaluated from the bounds of  $[x]$  and  $[y]$  for the addition, subtraction, multiplication, and division when  $0 \notin [y]$ .

The *range* of a continuous function  $f : \mathcal{D} \subset \mathbb{R} \rightarrow \mathbb{R}$  over an interval  $[x] \subset \mathcal{D}$

$$f([x]) = \{f(x) \mid x \in [x]\} \quad (13)$$

is again easily obtained when  $f$  is monotonic from evaluations involving the bounds of  $[x]$ . For elementary non-monotonic functions, such as all trigonometric functions, simple algorithms may be put at work to evaluate (13). In the general case, however,  $f([x])$  cannot be computed, which makes the concept of inclusion function particularly important.

An *inclusion function*  $[f]([x])$  of a function  $f(x)$  is such that

$$\forall [x] \subset \mathcal{D}, f([x]) \subset [f]([x]), \quad (14)$$

with  $[f]([x])$  an interval. A *minimal* inclusion function provides the smallest interval containing  $f([x])$  for all  $[x] \subset \mathcal{D}$ .

Various types of inclusion functions are available. The simplest one is the *natural* inclusion function  $[f_n]([x])$ , obtained by replacing, in the formal expression of  $f$ , all occurrences of the real variable  $x$  by its interval counterpart  $[x]$  and by performing all operations and elementary function evaluations on intervals.

Inclusion functions in general provide quite coarse outer-approximations of  $f([x])$  when they are many occurrences of  $[x]$  in their formal expression. Pessimism decreases when the width of the interval argument decreases, however.

All these notions are extended to interval vectors or *boxes*, which are Cartesian product of intervals, and to vector-valued functions. See [9] for more details.

#### C. SIVIA

Consider the problem of characterizing the set

$$\mathbb{X} = [\mathbf{x}] \cap \mathbf{f}^{-1}(\mathbb{Y}), \quad (15)$$

where  $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbb{Y} \subset \mathbb{R}^m$ , and  $[\mathbf{x}] \subset \mathcal{D}$  is some initial search box for  $\mathbb{X}$ . The aim of the *Set Inverter Via Interval Analysis* (SIVIA) [10] is to provide an inner approximation  $\underline{\mathbb{X}}$  and an outer approximation  $\overline{\mathbb{X}}$  of  $\mathbb{X}$ , represented by *subpavings*, i.e., unions of non-overlapping boxes. The distance between  $\underline{\mathbb{X}}$  and  $\overline{\mathbb{X}}$  is indicative of the quality of the approximation of  $\mathbb{X}$ . SIVIA requires an inclusion function  $[f]$  to be available for  $\mathbf{f}$ .

SIVIA iteratively partitions the box  $[\mathbf{x}]$  into subboxes on which the following tests are applied. Consider a given subbox  $[\tilde{\mathbf{x}}]$  of  $[\mathbf{x}]$ . If  $[f]([\tilde{\mathbf{x}}]) \subset \mathbb{Y}$  then (14) implies that  $\mathbf{f}([\tilde{\mathbf{x}}]) \subset \mathbb{Y}$  and thus that  $[\tilde{\mathbf{x}}] \subset \mathbb{X}$ . In this case,  $[\tilde{\mathbf{x}}]$  is stored

in  $\underline{\mathbb{X}}$  and in  $\overline{\mathbb{X}}$ . If  $[f]([\tilde{\mathbf{x}}]) \cap \mathbb{Y} = \emptyset$  then (14) implies that  $\mathbf{f}([\tilde{\mathbf{x}}]) \cap \mathbb{Y} = \emptyset$  and thus that  $[\tilde{\mathbf{x}}] \cap \mathbb{X} = \emptyset$ . In this case,  $[\tilde{\mathbf{x}}]$  is not considered any further. If none of the two previous tests is true,  $[\tilde{\mathbf{x}}]$  is undetermined. In this case, if  $[\tilde{\mathbf{x}}]$  is large enough, *i.e.*, if the largest width of its component intervals is larger than some precision parameter  $\varepsilon$ ,  $[\tilde{\mathbf{x}}]$  is bisected into  $[\tilde{\mathbf{x}}_1]$  and  $[\tilde{\mathbf{x}}_2]$  on which the previous tests are applied again. If  $[\tilde{\mathbf{x}}]$  is too small to be bisected, it is stored in  $\overline{\mathbb{X}}$ .

The efficiency of SIVIA is conditioned by the accuracy of the inclusion function available for  $\mathbf{f}$ .

To address the set inversion problem introduced in Section III-A with SIVIA, an inclusion function for  $\tau$  is necessary, which is based on inclusion functions for the  $\tau_i$ 's.

#### D. Contractors for guaranteed characterization

Large enough undetermined boxes need to be bisected by SIVIA. Indetermination often results from range over-estimation by inclusion functions. As a consequence, boxes have to be bisected many times to allow one to conclude on the position of the resulting boxes with respect to  $\mathbb{X}$ . This may entail intractable computational complexity, even for a moderate dimension of  $\mathbf{p}$ .

Contractors [9] partly address this issue. A contractor  $\mathcal{C}_{\mathbf{f},\mathbb{Y}}$  associated with the generic set-inversion problem (15) is a function taking a box  $[\mathbf{x}]$  as input and returning a box

$$\mathcal{C}_{\mathbf{f},\mathbb{Y}}([\mathbf{x}]) \subset [\mathbf{x}] \quad (16)$$

such that

$$[\mathbf{x}] \cap \mathbb{X} = \mathcal{C}_{\mathbf{f},\mathbb{Y}}([\mathbf{x}]) \cap \mathbb{X}, \quad (17)$$

so no part of  $\mathbb{X}$  in  $[\mathbf{x}]$  is lost. It allows parts of the candidate box  $[\mathbf{x}]$  that do not belong to  $\mathbb{X}$  to be eliminated, without the need to perform any bisection. Various types of contractors have been proposed in the literature, *e.g.*, the contractors by interval constraint propagation, by parallel linearization, the Newton contractor, the Krawczyk contractor, *etc.*

Here, the role of  $\mathbf{x}$  is taken by  $\mathbf{p}$ . The fact that the function  $\tau$  introduced in (11) is not differentiable forbids the use of most classic contractors, so specific contractors are needed. The new contractor proposed in this paper is implemented in two steps. It assumes that the functions  $z_i - z_0$  involved in (8) and (9) are differentiable. First, a set of  $m$  possibly overlapping subboxes of  $[\mathbf{p}]$  are built, trying to remove all values of  $\mathbf{p}$  from  $[\mathbf{p}]$  such that  $z_i(\mathbf{p}) - z_0(\mathbf{p}) < 0$ ,  $i = 1, \dots, m-1$ , see Section III-D.1. Second, the union of all non-empty intersections of at least  $q$  of these boxes is computed to get a possibly contracted box, see Section III-D.2.

1) *Box contraction using the  $(z_i - z_0)$ 's:* To build a contractor  $\mathcal{C}_{z_i - z_0, [0, \infty[}$  for the set of all values of  $\mathbf{p} \in [\mathbf{p}]$  such that  $z_i(\mathbf{p}) - z_0(\mathbf{p}) \geq 0$ , we take advantage of the fact that the functions  $\mathbf{s}_i(\mathbf{p})$ ,  $i = 0, \dots, m-1$  are affine in  $\mathbf{p}$  to reduce the number of occurrences of  $\mathbf{p}$  in their formal expression, and thus the pessimism of the corresponding

inclusion functions. Equation (5) can be rewritten as

$$\mathbf{s}_0(\mathbf{p}) = \sum_{t=1}^n y_t \varphi_t - \left( \sum_{t=1}^n \varphi_t \varphi_t^T \right) \mathbf{p} \quad (18)$$

$$= \mathbf{b}_0 - \mathbf{A}_0 \mathbf{p} \quad (19)$$

with  $\mathbf{b}_0 = \sum_{t=1}^n y_t \varphi_t$  and  $\mathbf{A}_0 = \sum_{t=1}^n \varphi_t \varphi_t^T$ . Similarly, (6) may be rewritten as

$$\mathbf{s}_i(\mathbf{p}) = \mathbf{b}_i - \mathbf{A}_i \mathbf{p} \quad (20)$$

with  $\mathbf{b}_i = \sum_{t=1}^n \alpha_{i,t} y_t \varphi_t$  and  $\mathbf{A}_i = \sum_{t=1}^n \alpha_{i,t} \varphi_t \varphi_t^T$ .

Using (19), (20), and the fact that the  $\mathbf{A}_i$ 's are symmetric, one gets

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = (\mathbf{b}_i - \mathbf{A}_i \mathbf{p})^T (\mathbf{b}_i - \mathbf{A}_i \mathbf{p}) - (\mathbf{b}_0 - \mathbf{A}_0 \mathbf{p})^T (\mathbf{b}_0 - \mathbf{A}_0 \mathbf{p}) \quad (21)$$

$$= \mathbf{p}^T (\mathbf{A}_i^2 - \mathbf{A}_0^2) \mathbf{p} - 2(\mathbf{b}_i^T \mathbf{A}_i - \mathbf{b}_0^T \mathbf{A}_0) \mathbf{p} + (\mathbf{b}_i^T \mathbf{b}_i - \mathbf{b}_0^T \mathbf{b}_0). \quad (22)$$

The matrices  $\mathbf{A}_i^2 - \mathbf{A}_0^2$  are symmetric and they may thus be diagonalized as  $\mathbf{A}_i^2 - \mathbf{A}_0^2 = \mathbf{U}_i^T \mathbf{D}_i \mathbf{U}_i$ , where  $\mathbf{U}_i$  is an orthonormal matrix (*i.e.*, such that  $\mathbf{U}_i^T = \mathbf{U}_i^{-1}$ ), and  $\mathbf{D}_i = \text{diag}(d_{i,1}, \dots, d_{i,n_p})$  is a diagonal matrix. Using the change of variables  $\boldsymbol{\pi} = \mathbf{U}_i \mathbf{p}$ , (22) becomes

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \boldsymbol{\pi}^T \mathbf{D}_i \boldsymbol{\pi} - 2\beta_i^T \boldsymbol{\pi} + \gamma_i, \quad (23)$$

with  $\beta_i^T = (\mathbf{b}_i^T \mathbf{A}_i - \mathbf{b}_0^T \mathbf{A}_0) \mathbf{U}_i^T$  and  $\gamma_i = \mathbf{b}_i^T \mathbf{b}_i - \mathbf{b}_0^T \mathbf{b}_0$ . Then, assuming that  $d_{i,j} \neq 0$  for  $j = 1, \dots, n_p$ , (23) can be rewritten as

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \sum_{j=1}^{n_p} d_{i,j} \left( \pi_j - \frac{\beta_{i,j}}{d_{i,j}} \right)^2 + \gamma_i - \sum_{j=1}^{n_p} \frac{\beta_{i,j}^2}{d_{i,j}}. \quad (24)$$

Consider  $[\boldsymbol{\pi}] = \mathbf{U}_i [\mathbf{p}]$ . A contractor for  $[\pi_j]$  is obtained from (24) as follows

$$[\pi'_j] = [\pi_j] \cap \left( \pm \left( \left( ([z_i]([\mathbf{p}]) - [z_0]([\mathbf{p}])) \cap [0, \infty[ \right) - \sum_{\substack{k=1 \\ k \neq j}}^{n_p} d_{i,k} \left( [\pi_k] - \frac{\beta_{i,k}}{d_{i,k}} \right)^2 - \gamma_j + \sum_{k=1}^{n_p} \frac{\beta_{i,k}^2}{d_{i,k}} \right) / d_{i,j} \right)^{\frac{1}{2}} + \frac{\beta_{i,j}}{d_{i,j}} \right). \quad (25)$$

From (25), the contractor  $\mathcal{C}_{z_i - z_0, [0, \infty[}$  for  $[\mathbf{p}]$  is obtained as

$$[\mathbf{p}'_i] = \mathcal{C}_{z_i - z_0, [0, \infty[}([\mathbf{p}]) = [\mathbf{p}] \cap (\mathbf{U}_i^T [\boldsymbol{\pi}'_i]). \quad (26)$$

When  $n$  is large enough and provided that the  $\varphi_t$ s have been well designed, it is very unlikely that  $\mathbf{A}_i^2 - \mathbf{A}_0^2$  is rank deficient. May this occur, (24) and (25) would have to be rewritten distinguishing the zero and nonzero  $d_{i,j}$ s.

*Proposition 1:* Assuming that  $d_{i,j} \neq 0$  for  $j = 1, \dots, n_p$ , for all  $[\mathbf{p}'_i]$ ,  $i = 1, \dots, m-1$ , built using (25) and (26), one has  $[\mathbf{p}'_i] \subset [\mathbf{p}]$  and

$$[\mathbf{p}'_i] \cap (z_i - z_0)^{-1}([0, \infty[) = [\mathbf{p}] \cap (z_i - z_0)^{-1}([0, \infty[). \quad (27)$$

*Proof:*  $[\mathbf{p}'_i] \subset [\mathbf{p}]$  is true by construction. To prove (27), it remains to prove that  $[\mathbf{p}] \cap (z_i - z_0)^{-1}([0, \infty[) \subset$

$[\mathbf{p}'_i] \cap (z_i - z_0)^{-1}([0, \infty[)$ . Consider  $\mathbf{p}^0 \in [\mathbf{p}] \cap (z_i - z_0)^{-1}([0, \infty[)$  and  $\boldsymbol{\pi}^0 = \mathbf{U}_i \mathbf{p}^0 \in [\boldsymbol{\pi}]$ . To prove that  $\mathbf{p}^0 \in [\mathbf{p}'_i] \cap (z_i - z_0)^{-1}([0, \infty[)$ , it suffices to prove that  $\boldsymbol{\pi}^0 \in [\boldsymbol{\pi}'_i]$ . By definition of  $\mathbf{p}^0$ , one has

$$\sum_{j=1}^{n_p} d_{i,j} \left( \pi_j^0 - \frac{\beta_{i,j}}{d_{i,j}} \right)^2 + \gamma_i - \sum_{j=1}^{n_p} \frac{\beta_{i,j}^2}{d_{i,j}} = z_i(\mathbf{p}^0) - z_0(\mathbf{p}^0) \geq 0. \quad (28)$$

Thus, since  $\boldsymbol{\pi}^0 \in [\boldsymbol{\pi}]$ , after some manipulations of (28), one gets

$$\begin{aligned} \pi_j^0 \in [\pi'_j] \cap \left( \pm \left( \left( ([z_i]([\mathbf{p}]) - [z_0]([\mathbf{p}])) \cap [0, \infty[ \right. \right. \right. \\ \left. \left. \left. - \sum_{\substack{k=1 \\ k \neq j}}^{n_p} d_{i,k} \left( [\pi_k] - \frac{\beta_{i,k}}{d_{i,k}} \right)^2 - \gamma_j + \sum_{k=1}^{n_p} \frac{\beta_{i,k}^2}{d_{i,k}} \right) / d_{i,j} \right)^{\frac{1}{2}} + \frac{\beta_{i,j}}{d_{i,j}} \right) \\ \in [\pi'_j], \end{aligned}$$

which completes the proof.  $\blacksquare$

2) *Building a  $q$ -relaxed intersection:* During the second step, the contractor builds a box  $[\mathbf{p}']$  enclosing the  $q$ -relaxed intersection  $\mathcal{P}$  [7], [8], [11] of the boxes in  $\mathcal{L} = \{[\mathbf{p}'_1], \dots, [\mathbf{p}'_{m-1}]\}$ , i.e., the union of all intersections of at least  $q$  boxes in  $\mathcal{L}$

$$\mathcal{P} = \bigcap_{j \in \{1, \dots, m-1\}}^q [\mathbf{p}'_j] = \bigcup_{\substack{J \subset \{1, \dots, m-1\} \\ \text{card}(J) \geq q}} \bigcap_{j \in J} [\mathbf{p}'_j] \quad (29)$$

and satisfying

$$\mathcal{P} \subset [\mathbf{p}'] \subset [\mathbf{p}]. \quad (30)$$

*Proposition 2:* For any box  $[\mathbf{p}']$ , satisfying (30), one has

$$\Sigma_q \cap [\mathbf{p}'] = \Sigma_q \cap [\mathbf{p}], \quad (31)$$

with  $\Sigma_q$  as defined in (8).

*Proof:* Assume that there exists  $\mathbf{p}_0 \in [\mathbf{p}]$  such that  $\mathbf{p}_0 \in \Sigma_q \cap [\mathbf{p}]$  but  $\mathbf{p}_0 \notin \Sigma_q \cap [\mathbf{p}']$ . Since  $\mathbf{p}_0 \in \Sigma_q \cap [\mathbf{p}]$ ,  $\mathbf{p}_0 \in \Sigma_q$ . According to (8),  $\sum_{i=1}^{m-1} \tau_i(\mathbf{p}_0) \geq q$ . There are thus at least  $q$  functions  $\tau_i$  such that  $\tau_i(\mathbf{p}_0) \geq 1$ . Assume, without loss of generality, that  $\tau_1(\mathbf{p}_0) \geq 1, \dots, \tau_q(\mathbf{p}_0) \geq 1$ . Since  $\tau_i(\mathbf{p}_0) \geq 1$ ,  $i = 1, \dots, q$ , by definition of  $\mathcal{C}_{z_i - z_0, [0, \infty[}$ , one has  $\mathbf{p}_0 \in [\mathbf{p}'_i]$ ,  $i = 1, \dots, q$  and  $\mathbf{p}_0 \in \bigcap_{i=1, \dots, q} [\mathbf{p}'_i]$ . By definition of  $\mathcal{P}$  and  $[\mathbf{p}']$ ,  $\mathbf{p}_0 \in \bigcap_{i=1, \dots, q} [\mathbf{p}'_i] \subset \mathcal{P} \subset [\mathbf{p}']$ , which contradicts the initial assumption.  $\blacksquare$

a) *Evaluating the  $q$ -relaxed intersection:* Algorithm 1 formalizes a computation carried out on an example in [8]. It aims at building an outer approximating interval of the  $q$ -relaxed intersection of  $m$  scalar intervals. The extension to boxes is obtained by applying Algorithm 1 componentwise.

Consider a list  $\mathcal{L} = \{[p_1], \dots, [p_{m-1}]\}$  of  $m-1$  scalar intervals. Algorithm 1 builds the smallest interval containing the union of all intersections of  $q$  intervals with a complexity  $O(m \log m)$ . This is the smallest interval containing  $\mathcal{P}$  as defined by (29) in the scalar case. At Steps 4 and 8 of Algorithm 1,  $(p \in [p_j]) = 1$  if  $p \in [p_j]$  and  $(p \in [p_j]) = 0$  otherwise.

**Algorithm 1**  $[p] = q$ -relaxed intersection  $([p_1], \dots, [p_{m-1}])$

```

1   $[p] = \emptyset$ ;
2  Reindex the boxes  $[p_i]$  in such a way that
    $p_1 \leq p_2 \leq \dots \leq p_{m-1}$ ;
3  For  $i = q$  to  $m-1$ 
4     if  $\sum_{j=1}^{m-1} (p_j \in [p_i]) \geq q$ 
5         $p = p_i$ ; break;
6  Reindex the boxes  $[p_i]$  in such a way that
    $\bar{p}_1 \geq \bar{p}_2 \geq \dots \geq \bar{p}_{m-1}$ 
7  For  $i = q$  to  $m-1$ 
8     if  $\sum_{j=1}^{m-1} (\bar{p}_j \in [p_i]) \geq q$ 
9         $\bar{p} = \bar{p}_i$ ; break;
```

When  $\mathcal{L}$  is a list of boxes of the same dimensions, one may simply apply Algorithm 1 component by component.

## IV. EXAMPLES

Consider the system

$$y_t = y_t^m(\mathbf{p}) + w_t, \quad (32)$$

with the FIR model

$$y_t^m(\mathbf{p}) = \sum_{i=0}^{n_a-1} a_i u_{t-i}, \quad (33)$$

where  $\mathbf{p} = (a_0, \dots, a_{n_a-1})^T$  and  $u_n = 0$  for  $n \leq 0$ . For  $t = 1, \dots, n$ , the  $w_t$ s are independent and identically distributed (iid) noise samples. In linear regression form, (32) becomes

$$y_t = \boldsymbol{\varphi}_t^T \mathbf{p}^* + w_t \quad (34)$$

with  $\boldsymbol{\varphi}_t = (u_t, \dots, u_{t-n_a+1})^T$  and  $\mathbf{p}^* = (a_0^*, \dots, a_{n_a-1}^*)^T$ .

### A. Lower-dimensional model

When the dimension of  $\mathbf{p}$  is small, the characterization of  $\Sigma_q$  introduced in (8) can be addressed using SIVIA. For that purpose, the following inclusion functions for the  $\tau_i$ 's are introduced

$$[\tau_i]([\mathbf{p}]) = \begin{cases} 1 & \text{if } \inf([z_i - z_0]([\mathbf{p}])) \geq 0, \\ 0 & \text{if } \sup([z_i - z_0]([\mathbf{p}])) < 0, \\ [0, 1] & \text{else,} \end{cases} \quad (35)$$

where  $[z_i - z_0]([\mathbf{p}])$  is an inclusion function for the difference between  $z_i$  and  $z_0$ .

1) *Laplacian noise:* Data are generated for the actual system parameters  $a_0^* = 0.2$ ,  $a_1^* = 0.3$ , and  $a_2^* = 0.4$  considering: (a) a filtered Gaussian input

$$u_t = \alpha u_{t-1} + v_t, \quad (36)$$

with  $\alpha = 0.2$  and  $v_t$  iid zero-mean Gaussian with variance  $\sigma_v^2 = 0.65$  and (b) a random iid sequence of  $\pm 1$ , which is the D-optimal input under the constraint that the input has to remain in  $[-1, 1]$ . In both cases, the noise samples  $w_t$  are zero-mean Laplacian with standard deviation  $\sigma_w$  tuned to get a signal-to-noise ratio (SNR) of 15 dB. We choose  $n = 1024$ ,  $m = 255$ , and  $q = 13$ . Our aim is thus to characterize a 95%

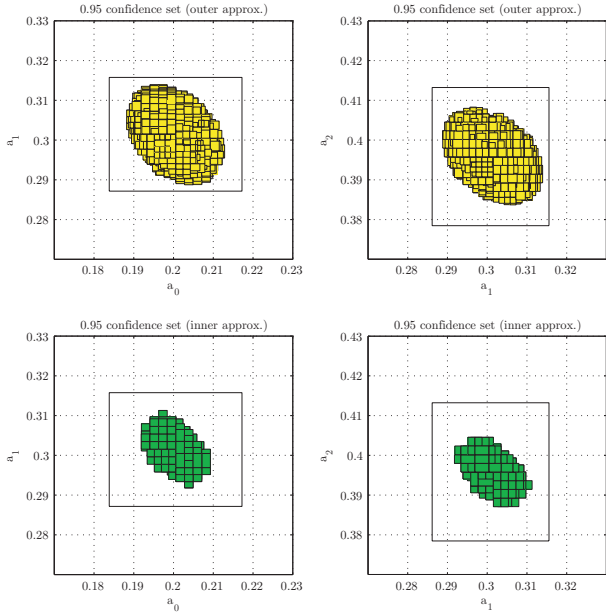


Fig. 1. Projections on the  $(a_0, a_1)$ -plane (left) and  $(a_1, a_2)$ -plane (right) of the subpavings of the search space obtained using SPS in the FIR case, with a filtered Gaussian input; the 95 % confidence region is contained in the union of yellow and green boxes; the box containing the subpavings is obtained after a single application of the contractor of Section III-D

confidence region. Undetermined boxes are bisected when their maximum width is larger than  $\varepsilon = 2.5 \times 10^{-3}$ .

Figures 1 and 2 show projections of the subpavings of the search box  $\mathbb{P} = [-10^4, 10^4]^3$ . The green boxes are contained in  $\Sigma_q$  and the union of yellow and green boxes form an outer approximation of  $\Sigma_q$ . The box containing the subpavings results from a single application of the contractor of Section III-D taking  $\mathbb{P}$  as input. This box is very close to the outer-approximation of  $\Sigma_q$ , showing the efficiency of the contractor of Section III-D. It does not correspond to the smallest box containing  $\Sigma_q$  due to the fact that for each  $\tau_i$ , a reduced-size box  $[\pi_i]$  is evaluated in a specific coordinate system. The inverse change of variable involves a rotation of  $[\pi_i]$  and the resulting box  $[\mathbf{p}_i]$  is the smallest box containing the rotated  $[\pi_i]$ . The wrapping effect introduces some pessimism, which propagates to the result of the  $q$ -relaxed intersection.

The D-optimal input provides better results in terms of size of uncertainty. The total volume of the green boxes and of the green and yellow boxes is respectively  $2.0 \times 10^{-6}$  and  $6.6 \times 10^{-6}$  for the filtered Gaussian input, where it is  $1.25 \times 10^{-6}$  and  $3.6 \times 10^{-6}$  for the D-optimal input.

2) *Laplacian-Bernoulli-Laplacian noise*: A third set of data is generated with the same parameters as before, except for the measurement noise: 5 % of the noise samples were replaced by Laplacian noise samples with a standard deviation equal to  $10\sigma_w$ . This reduces the SNR to 9.5 dB. SIVIA is used again to perform the characterization of  $\Sigma_q$ . The resulting subpavings of  $\mathbb{P} = [-10^4, 10^4]^3$  are represented in Figure 3. Their volume has significantly increased:  $1.5 \times 10^{-5}$  for the green boxes and  $2.4 \times 10^{-5}$  for the green and yellow boxes. Nevertheless, even with such an impulsive noise, one is able to characterize an exact confidence region

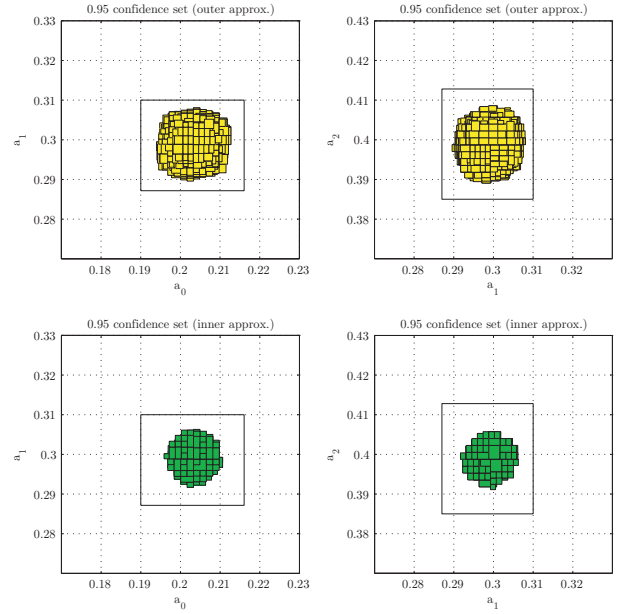


Fig. 2. Projections on the  $(a_0, a_1)$ -plane (left) and  $(a_1, a_2)$ -plane (right) of the subpavings of the search space obtained using SPS in the FIR case, with a D-optimal input; the 95 % confidence region is contained in the union of yellow and green boxes; the box containing the subpaving is obtained after a single application of the contractor of Section III-D

in a guaranteed way, considering only the measurements, and not taking into account the noise distribution.

## B. Higher-dimensional model

For models with a large number of parameters (typically more than 5 parameters), the computational complexity to get subpavings providing inner and outer-approximations of the confidence region is prohibitive. Nevertheless, it is still possible to obtain a box containing  $\Sigma_q$  in a guaranteed way via the contractor introduced in Section III-D.

To evaluate the performance of the proposed technique for a larger number of parameters, FIR models (33) with  $n_a = 20$  random parameters in  $[-2, 2]^{n_a}$  are generated. Then,  $n = 512, 1024, 2048, 4096,$  and  $8192$  noise-free data points are first generated applying to (33) sequences with the same characteristics as in Section IV-A.1. White Laplacian noise is then added to these data. The standard deviation of the noise is set up to get an SNR of 5 dB to 40 dB.

We choose  $m = 255$  and  $q = 13$ . Our aim is thus again to characterize a 95 % confidence region. The initial search box in the parameter space is taken as  $\mathbb{P} = [-10^4, 10^4]^{20}$ . The contractor of Section III-D is applied once to  $\mathbb{P}$  (iterated applications are useless). Figures 4 and 5 represent the width of the largest component of the resulting box as a function of the SNR and of the number of data points for the filtered Gaussian input (dotted line) and for the D-optimal input (solid line).

As for the previous examples, the D-optimal input provides a better estimation accuracy. On a log-log scale, maximum width seems linear in the SNR (see Figure 4) and in the number of samples (see Figure 5). In the latter case, the slope is about  $-1/2$ , consistent with what is observed when maximum-likelihood estimation is carried out assuming an

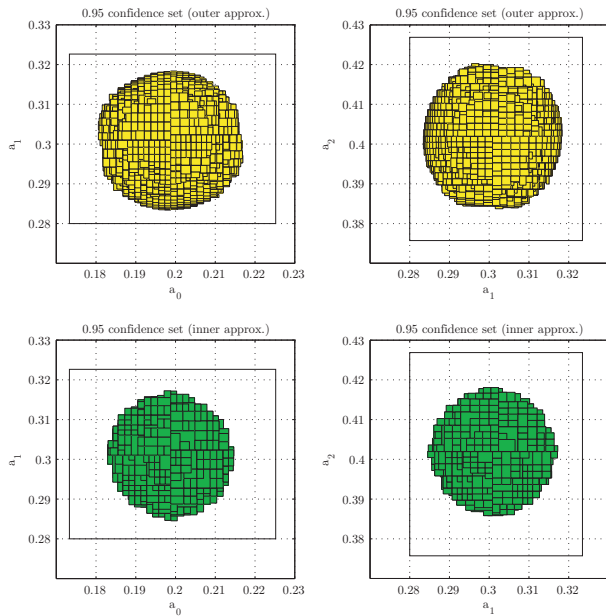


Fig. 3. Projections on the  $(a_0, a_1)$ -plane (left) and  $(a_1, a_2)$ -plane (right) of the subpaving of the search space obtained using SPS in the FIR case, D-optimal input signal, Laplacian-Bernoulli-Laplacian noise; the 95 % confidence region is contained in the union of yellow and green boxes; the box containing the subpaving is obtained after a single application of the contractor of Section III-D

additive Gaussian noise, although this hypothesis on the noise is neither true nor assumed here.

## V. CONCLUSIONS AND PERSPECTIVES

Interval analysis provides tools for evaluating guaranteed approximations of the exact confidence regions defined by SPS. Interval methods that rely on bisections, such as SIVIA, are limited to problems with a small number of parameters. The contractor proposed in this paper, on the other hand makes it possible to deal with FIR models with a large number of parameters, which is the rule for FIR model but usually a nightmare for interval methods. The results are spectacular, as the methods previously employed to characterize the confidence regions provided by SPS would have been completely unable to deal with a FIR model with

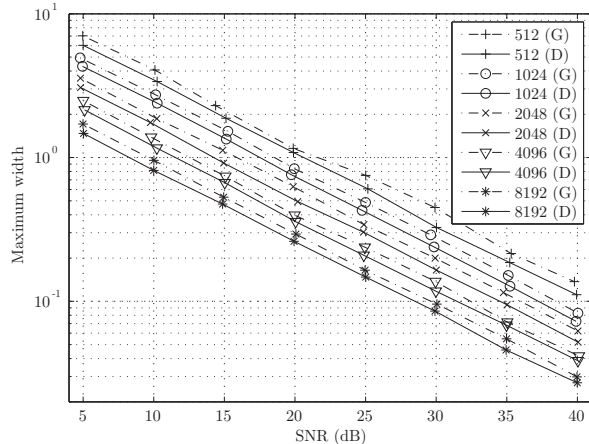


Fig. 4. Width of the largest component of the box resulting from a single application of the contractor of Section III-D as a function of the SNR and of the number of data points for the filtered Gaussian input (dotted) and for the D-optimal input (plain).

20 parameters, let alone to provide any guarantee as to their results. Research on efficient contractors is also the key to dealing with large-scale models with the LSCR method.

## REFERENCES

- [1] M. C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005.
- [2] M. C. Campi and E. Weyer. Non-asymptotic confidence sets for the parameters of linear transfer functions. *IEEE trans. Autom. Control*, 55(12):2708–2720, 2010.
- [3] B. C. Csáji, M. C. Campi, and E. Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proc. IFAC SYSID*, pages 227–232, Brussels, Belgium, 2012.
- [4] M. Dalai, E. Weyer, and M. C. Campi. Parameter identification for nonlinear systems: Guaranteed confidence regions through LSCR. *Automatica*, 43:1418 – 1425, 2007.
- [5] O. N. Granichin. The nonasymptotic confidence set for parameters of a linear control object under an arbitrary external disturbance. *Automation and Remote Control*, 73(1):20–30, 2012.
- [6] P. Hougaard. Saddlepoint approximations for curved exponential families. *Statistics and Probability Letters*, 3(3):161–166, 1985.
- [7] L. Jaulin. Robust set membership state estimation; application to underwater robotics. *Automatica*, 45(1):202–206, 2009.
- [8] L. Jaulin. Set-membership localization with probabilistic errors. *Robotics and Autonomous Systems*, 59(6):489–495, 2011.
- [9] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis*. Springer-Verlag, London, 2001.
- [10] L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4):1053–1064, 1993.
- [11] L. Jaulin and E. Walter. Guaranteed robust nonlinear minimax estimation. *IEEE Trans. Autom. Cont.*, 47(11):1857–1864, 2002.
- [12] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.
- [13] M. Kieffer and E. Walter. Guaranteed characterization of exact non-asymptotic confidence regions in nonlinear parameter estimation. In *Proc. IFAC NOLCOS*, Toulouse, France, pages 56–61, 2013.
- [14] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter, editors. *Bounding Approaches to System Identification*. Plenum Press, New York, NY, 1996.
- [15] R. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [16] A. Pázman. Small-sample distribution properties of nonlinear regression estimators (a geometric approach) (with discussion). *Statistics*, 21(3):323–367, 1990.
- [17] A. Pázman and L. Pronzato. Densities of nonlinear functions of the nonlinear least-squares estimator. In M. Blanke and T. Söderstrom, editors, *Prep. IFAC SYSID*, vol. 1, pages 163–168, Copenhagen, 1994.
- [18] I. M. Skovgaard. Large deviation approximations for maximum likelihood estimators. *Prob. Math. Stat.*, 6(2):89–107, 1985.
- [19] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Springer-Verlag, London, 1997.

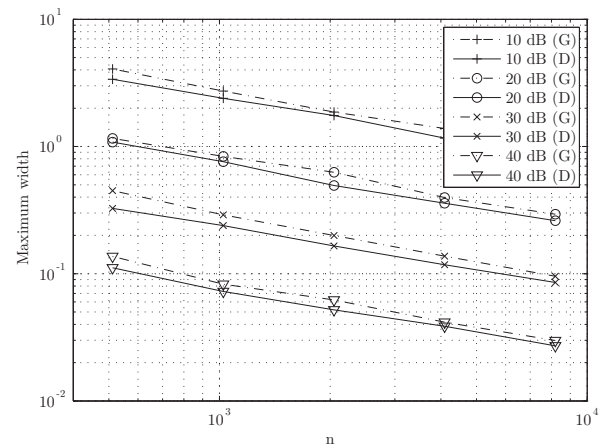


Fig. 5. Width of the largest component of the box resulting from a single application of the contractor of Section III-D as a function of the number of data points and of the SNR for the filtered Gaussian input (dotted) and for the D-optimal input (plain).