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, 2013, <10.1109/cdc.2013.6760681 >. <hal-00879946>

HAL Id: hal-00879946
https://hal-supelec.archives-ouvertes.fr/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.

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

\[
y_t = \varphi_t^T \mathbf{p}^* + w_t, t = 1, \ldots, n, \quad (3)
\]

with \( \varphi_t \) 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

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, 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 bld Saint-Michel, 75005 Paris
the solution to the normal equations

$$\sum_{i=1}^{n} \varphi_i \left(y_i - \varphi_i^\top \mathbf{p}\right) = 0.$$  (4)

For a generic \( \mathbf{p} \), define

$$s_0 \left( \mathbf{p} \right) = \sum_{i=1}^{n} \varphi_i \left(y_i - \varphi_i^\top \mathbf{p}\right),$$  (5)

and the sign-perturbed sums

$$s_i \left( \mathbf{p} \right) = \sum_{i=1}^{n} \alpha_{i,t} \varphi_i \left(y_i - \varphi_i^\top \mathbf{p}\right),$$  (6)

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

$$z_i \left( \mathbf{p} \right) = \| s_i \left( \mathbf{p} \right) \|_2^2, \quad i = 0, \ldots, m - 1.$$  (7)

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

\( \Sigma_q \) may be defined more formally as

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

where for \( i = 1, \ldots, m - 1 \),

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

This is justified by the fact that if \( \sum_{i=1}^{m-1} \tau_i \left( \mathbf{p} \right) \geq q \), then \( \tau_i \left( \mathbf{p} \right) = 1 \) for at least \( q \) out of the \( m - 1 \) functions \( \tau_i \left( \mathbf{p} \right) \).

As a consequence, there are at least \( q \) functions \( z_i \left( \mathbf{p} \right) \) such that \( z_i \left( \mathbf{p} \right) > z_0 \left( \mathbf{p} \right) \) and \( z_0 \left( \mathbf{p} \right) \) is not among the \( q \) largest values of \( \left(z_i \left( \mathbf{p} \right)\right)_{i=0}^{m-1} \).

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

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

with

$$\tau \left( \mathbf{p} \right) = \sum_{i=1}^{m-1} \tau_i \left( \mathbf{p} \right),$$  (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 \( [x] = [x, \overline{x}] \) of \( \mathbb{R} \) and extends all arithmetic operations and elementary functions on real numbers to these intervals.

For arithmetic operations,

$$[x] \circ [y] = \{ x \circ y \mid x \in [x], \ y \in [y] \},$$  (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 \( f : \mathcal{D} \subset \mathbb{R} \to \mathbb{R} \) over an interval \( [x] \subset \mathcal{D} \)

$$f \left([x]\right) = \{ f \left(x\right) \mid x \in [x] \}$$  (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, \( f \left([x]\right) \) cannot be computed, which makes the concept of inclusion function particularly important.

An inclusion function \( f \left([x]\right) \) of a function \( f \left(x\right) \) is such that

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

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

Various types of inclusion functions are available. The simplest one is the natural inclusion function \( \left[f_0\right] \left([x]\right) \), 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 \( f \left([x]\right) \) 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

$$\Xi = [x] \cap f^{-1} \left( \mathbb{Y} \right),$$  (15)

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

SIVIA iteratively partitions the box \( [x] \) into subboxes on which the following tests are applied. Consider a given subbox \( [\hat{x}] \) of \( [x] \). If \( [f] \left([\hat{x}]\right) \subset \mathbb{Y} \) then (14) implies that \( f \left([\hat{x}]\right) \subset \mathbb{Y} \) and thus that \( [\hat{x}] \subset \Xi \). In this case, \( [\hat{x}] \) is stored
in $\mathbb{X}$ and in $\overline{\mathbb{X}}$. If $[f](\overline{[x]}) \cap \overline{Y} = \emptyset$ then (14) implies that $f([x]) \cap Y = \emptyset$ and thus that $[x] \cap \mathbb{X} = \emptyset$. In this case, $[x]$ is not considered any further. If none of the two previous tests is true, $[x]$ is undetermined. In this case, if $[x]$ is large enough, i.e., if the largest width of its component intervals is larger than some precision parameter $\varepsilon$, $[x]$ is bisected into $[x_1]$ and $[x_2]$ on which the previous tests are applied again. If $[x]$ 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 $f$.

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 overestimation 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 $p$.

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

$$C_{t, Y}([x]) \subset [x]$$

such that

$$[x] \cap \mathbb{X} = C_{t, Y}([x]) \cap \mathbb{X},$$

so no part of $\mathbb{X}$ in $[x]$ is lost. It allows parts of the candidate box $[x]$ 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 $x$ is taken by $p$. 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 $[p]$ are built, trying to remove all values of $p$ from $[p]$ such that $z_i(p) - z_0(p) < 0$, $i = 1, \ldots, m - 1$, 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 $C_{z_i-z_0, [0, \infty]}$ for the set of all values of $p \in [p]$ such that $z_i(p) - z_0(p) \geq 0$, we take advantage of the fact that the functions $s_i(p)$, $i = 0, \ldots, m - 1$ are affine in $p$ to reduce the number of occurrences of $p$ in their formal expression, and thus the pessimism of the corresponding inclusion functions. Equation (5) can be rewritten as

$$s_0(p) = \sum_{t=1}^{n} y_t \varphi_t - \left( \sum_{t=1}^{n} \varphi_t \varphi_t^T \right) p$$  \hspace{1cm} (18)

$$= b_0 - A_0 p$$  \hspace{1cm} (19)

with $b_0 = \sum_{i=1}^{n} y_i \varphi_i$ and $A_0 = \sum_{t=1}^{n} \varphi_t \varphi_t^T$. Similarly, (6) may be rewritten as

$$s_t(p) = b_i - A_i p$$  \hspace{1cm} (20)

with $b_i = \sum_{t=1}^{n} \alpha_t \varphi_i \varphi_i^T$ and $A_i = \sum_{t=1}^{n} \alpha_t \varphi_i \varphi_i^T$.

Using (19), (20), and the fact that the $A_i$’s are symmetric, one gets

$$z_i(p) - z_0(p) (b_i - A_i p)^T (b_i - A_i p)$$

$$- (b_0 - A_0 p)^T (b_0 - A_0 p)$$

$$= p^T \left( A_i^2 - A_0^2 \right) p - 2 (b_i^T A_i - b_0^T A_0) p$$

$$+ (b_i^T b_i - b_0^T b_0).$$

(21)

The matrices $A_i^2 - A_0^2$ are symmetric and they may thus be diagonalized as $A_i^2 - A_0^2 = U_i \Omega_i U_i^T$, where $U_i$ is an orthonormal matrix (i.e., such that $U_i^T U_i = I$), and $D_i = \text{diag} (\{d_{i,1}, \ldots, d_{i,n_p}\})$ is a diagonal matrix. Using the change of variables $\pi = U_i p$, (22) becomes

$$z_i(p) - z_0(p) = \pi^T D_i \pi - 2 \beta_i^T \pi + \gamma_i$$  \hspace{1cm} (23)

with $\beta_i^T = (b_i^T A_i - b_0^T A_0) U_i^T$ and $\gamma_i = b_i^T b_i - b_0^T b_0$. Then, assuming that $d_{i,j} \neq 0$ for $j = 1, \ldots, n_p$, (23) can be rewritten as

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

(24)

Consider $\pi = U_i [p]$. A contractor for $[\pi_j]$ is obtained from (24) as follows

$$[\pi_j'] = [\pi_j] \cap \left( \pm \left( \left( [z_i] ([p]) - [z_0] ([p]) \right) \cap [0, \infty] \right) \right.$$  

$$- \sum_{k=1}^{n_p} d_{i,k} (|\pi_k| - \beta_{i,k})^2 / \gamma_j + \sum_{k=1}^{n_p} \beta_{i,k}^2 / d_{i,k} + \beta_{i,j}^2 / d_{i,j} \right)^{1/2}.$$  

(25)

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

$$[p_i'] = C_{z_i-z_0, [0, \infty]} \cap [p] \cap U_i^T (\pi') \cap [0, \infty].$$

(26)

When $n$ is large enough and provided that the $\varphi_i$’s have been well designed, it is very unlikely that $A_i^2 - A_0^2$ 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 $j = 1, \ldots, n_p$, for all $[p_i']$, $i = 1, \ldots, m - 1$, built using (25) and (26), one has $[p_i'] \subset [p]$ and

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

(27)

**Proof:** $[p_i'] \subset [p]$ is true by construction. To prove (27), it remains to prove that $[p] \cap (z_i - z_0)^{-1} ([0, \infty]) \subset$
Thus, since $\pi^0 \in [\pi]$, after some manipulations of (28), one gets

$$
\pi^0_j \in [\pi_j] \cap \left( \pm \left( \left( \left( [z_i]([p]) - [z_0]([p]) \right) \cap [0, \infty] \right) \right) \right)
$$

$$
- \sum_{k=1}^{n_p} d_{i,k} (\{\pi_k\}) - \gamma_j + \sum_{k=1}^{n_p} \frac{\beta_{i,k}^2}{d_{i,k}} \right) \right) \right) + \frac{\beta_{i,k}^2}{d_{i,k}} \right) \right)
$$

$$
\in [\pi_j'],
$$

which completes the proof.

2) Building a $q$-relaxed intersection: During the second step, the contractor builds a box $[p']$ enclosing the $q$-relaxed intersection $\mathcal{P}$ [7], [8], [11] of the boxes in $\mathcal{L} = \{[p_1], \ldots, [p_{m-1}]\}$, i.e., the union of all intersections of at least $q$ boxes in $\mathcal{L}$

$$
\mathcal{P} = \bigcup_{j \in \{1, \ldots, m-1\}}^{q} [p_j] = \bigcup_{J \subset \{1, \ldots, m-1\}, |J| \geq q} [p_j]
$$

and satisfying

$$
\mathcal{P} \subset [p'] \subset [p].
$$

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

$$
\Sigma_q \cap [p'] = \Sigma_q \cap [p],
$$

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

**Proof:** Assume that there exists $p_0 \in [p]$ such that $p_0 \in \Sigma_q \cap [p']$ but $p_0 \notin \Sigma_q \cap [p]$. Since $p_0 \in \Sigma_q \cap [p_0]$, $p_0 \in \Sigma_q$. According to (8), $\sum_{i=1}^{m-1} \tau_i(p_0) \geq q$. There are thus at least $q$ functions $\tau_i$ such that $\tau_i(p_0) \geq 1$. Assume, without loss of generality, that $\tau_i(p_0) \geq 1, \ldots, \tau_q(p_0) \geq 1$. Since $\tau_i(p_0) \geq 1, i = 1, \ldots, q$, by definition of $C_{z_i-z_0,[0,\infty]}$, one has $p_0 \in [p_i']$, $i = 1, \ldots, q$ and $p_0 \in \bigcap_{i=1}^{m-1} [p_i']$. By definition of $\mathcal{P}$ and $[p]'$, $p_0 \in \bigcap_{i=1}^{m-1} q \subset \mathcal{P} \subset [p']$, which contradicts the initial assumption.

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

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

$$
y_t = y_t^m(p) + w_t,
$$

with the FIR model

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

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

$$
y_t = \varphi_T^T p^* + w_t
$$

with $\varphi_t = (u_t, \ldots, u_{t-n_{s+1}})^T$ and $p^* = (a_0^*, \ldots, a_{n_p-1}^*)^T$.

#### A. Lower-dimensional model

When the dimension of $p$ 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

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

where $[z_i-z_0]([p])$ 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

$$
\alpha_t = \alpha u_{t-1} + v_t,
$$

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.

Fig. 2. 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, 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 $[-2, 2]^m$ are generated. Then, $n = 512, 1024, 2048, 4096, 8192$ 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 $P = [-10^4, 10^4]$. The contractor of Section III-D is applied once to $P$ (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 $\frac{1}{2}$, consistent with what is observed when maximum-likelihood estimation is carried out assuming an
The D-optimal input (plain).

of the number of data points for the filtered Gaussian input (dotted) and for

Fig. 3. Projections on the noise is neither true nor assumed here.

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 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


