HAL
open science

Guaranteed characterization of exact non-asymptotic confidence regions in nonlinear parameter estimation

Michel Kieffer, Eric Walter

To cite this version:

Michel Kieffer, Eric Walter. Guaranteed characterization of exact non-asymptotic confidence regions in nonlinear parameter estimation. 9th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2013), Sep 2013, Toulouse, France. pp.56-61, 10.3182/20130904-3-FR-2041.00019 . hal-00819488

HAL Id: hal-00819488

https://centralesupelec.hal.science/hal-00819488v1

Submitted on 1 May 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 non-asymptotic confidence regions in nonlinear parameter estimation

Michel Kieffer *,**,***, **Eric Walter** *,

* *L2S, CNRS, Supelec, Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France*
** *Partly on leave at LTCI, CNRS, Telecom ParisTech, 46 rue Barrault, 75013 Paris*
*** *Institut Universitaire de France, 103 bld Saint-Michel, 75005 Paris*

Abstract:
Recently, a new family of methods has been proposed for characterizing accuracy in nonlinear parameter estimation by Campi *et al.*. These methods make it possible to obtain exact, non-asymptotic confidence regions for the parameter estimates under relatively mild assumptions on the noise distribution, namely that the noise samples are independently and symmetrically distributed.
The numerical characterization of an exact confidence region with this new approach is far from being trivial, however. The aim of this paper is to show how interval analysis, which has been used for a guaranteed characterization of confidence regions for the parameter vector in other contexts, can contribute.

1. INTRODUCTION

When a vector $\mathbf{p}$ of parameters of some approximate mathematical model is estimated from a noisy data vector $\mathbf{y}$, this is usually via the minimization of some cost function $J(\mathbf{p})$, for instance

$$J(\mathbf{p}) = \|\mathbf{y} - \mathbf{y}_{\mathrm{m}}(\mathbf{p})\|_2^2 , \tag{1}$$

where $\mathbf{y}_{\mathrm{m}}(\mathbf{p})$ is the 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

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

This procedure is fraught with difficulties in the general nonlinear case. The parameters of the model may not be identifiable uniquely (*i.e.*, there may be several values of $\hat{\mathbf{p}}$ that yield exactly the same vector $\mathbf{y}_{\mathrm{m}}(\hat{\mathbf{p}})$, in which case there are several global minimizers of the cost function). The numerical algorithm used to compute $\hat{\mathbf{p}}$ may also get trapped at a parasitic local minimizer because of inadequate initialization.

Even if a single numerical vector $\hat{\mathbf{p}}$ is obtained and if $\mathbf{y}$ and $\mathbf{y}_{\mathrm{m}}(\hat{\mathbf{p}})$ 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. This is especially important if one wants to estimate physically meaningful parameters of some knowledge-based model in physics, chemistry, biology, *etc.*, or if decisions have to be taken on the basis of the numerical values of the model parameters to tune controllers or to detect faults, for instance.

⋆ This work has been partly supported by the ANR CPP.

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 LSCR approach (*Leave out Sign-dominant Correlation Regions*) recently developed by Campi *et al.* The paper is organized as follows. The LSCR approach is recalled in in Section 2. Section 3 shows how interval analysis can be used to characterize the exact confidence regions provided by LSCR in a global and guaranteed way. Examples are treated in Section 4, and conclusions drawn in Section 5.

2. CHARACTERIZING PARAMETER UNCERTAINTY WITH THE LSCR APPROACH

We assume in what follows that the system generating the data belongs to the model set, and denote its parameter vector by $\mathbf{p}^*$. This gives meaning to the concept of a true value for the parameter vector. It could be avoided by stating that we want to characterize the set of all values of $\mathbf{p}$ that are acceptable in a sense to be made precise.

In Campi and Weyer [2005], Dalai et al. [2007], Csáji et al. [2012], two new approaches named LSCR and SPS are described for the *exact* characterization of parameter uncertainty in *non-asymptotic* conditions. Their most striking feature is that they avoid 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 δ on the size of the acceptable errors is known. Both approaches require the noise samples to be independently distributed with distributions symmetric with respect to zero. LSCR is summarized in what follows, since it is not limited to linear models, contrary to SPS.

LSCR defines, *without any approximation*, regions $\mathbf{\Theta}$ to which the parameter vector $\mathbf{p}^*$ of the true system belongs with some specified probability.

Let $\varepsilon_t(\mathbf{p})$ be a prediction error, such that $\varepsilon_t(\mathbf{p}^*)$ is a realization of the noise corrupting the data at time t. The procedure for computing one such confidence region is as follows:

(1) Select two integers $r \geqslant 0$ and $q \geqslant 0$.
(2) For $t = 1 + r, \ldots, k + r = n$, compute
$$c^{\varepsilon}_{t-r,r}(\mathbf{p}) = \varepsilon_{t-r}(\mathbf{p})\,\varepsilon_t(\mathbf{p}), \tag{3}$$
(3) Compute
$$s^{\varepsilon}_{i,r}(\mathbf{p}) = \sum_{k \in \mathbb{I}_i} c^{\varepsilon}_{k,r}(\mathbf{p}), \; i = 1, \ldots, m, \tag{4}$$
where $\mathbb{I}_i$ is a subset of a set $\mathbb{I}$ of indexes and the collection $\mathbb{G}$ of these subsets $\mathbb{I}_i$, $i = 1, \ldots, m$, forms a group under the symmetric difference operation, *i.e.*, $(\mathbb{I}_i \cup \mathbb{I}_j) - (\mathbb{I}_i \cap \mathbb{I}_j) \in \mathbb{G}$.
(4) Find the set $\mathbf{\Theta}^{\varepsilon}_{r,q}$ such that *at least* q of the functions $s^{\varepsilon}_{i,r}(\mathbf{p})$ are *larger* than 0 and *at least* q are *smaller* than 0.

This set is such that the probability that $\mathbf{p}^*$ belongs to it can be computed *exactly*
$$\Pr\left(\mathbf{p}^* \in \mathbf{\Theta}^{\varepsilon}_{r,q}\right) = 1 - 2q/m. \tag{5}$$
The shape and size of $\mathbf{\Theta}^{\varepsilon}_{r,q}$ depend not only on the values given to q and r but also on the group $\mathbb{G}$ and its number of elements m. A procedure for generating a group of appropriate size is suggested in Gordon [1974].

The set $\mathbf{\Theta}^{\varepsilon}_{r,q}$ may be defined more formally as
$$\mathbf{\Theta}^{\varepsilon}_{r,q} = \mathbf{\Theta}^{\varepsilon,-}_{r,q} \cap \mathbf{\Theta}^{\varepsilon,+}_{r,q}, \tag{6}$$
with
$$\mathbf{\Theta}^{\varepsilon,-}_{r,q} = \left\{\mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau^{\varepsilon,-}_i(\mathbf{p}) \geqslant q\right\}, \tag{7}$$
$$\mathbf{\Theta}^{\varepsilon,+}_{r,q} = \left\{\mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau^{\varepsilon,+}_i(\mathbf{p}) \geqslant q\right\}, \tag{8}$$
where $\mathbb{P}$ is the prior domain for $\mathbf{p}$ and where
$$\tau^{\varepsilon,-}_i(\mathbf{p}) = \begin{cases} 1 & \text{if } -s^{\varepsilon}_{i,r}(\mathbf{p}) \geqslant 0, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$
and
$$\tau^{\varepsilon,+}_i(\mathbf{p}) = \begin{cases} 1 & \text{if } s^{\varepsilon}_{i,r}(\mathbf{p}) \geqslant 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$
The set $\mathbf{\Theta}^{\varepsilon,-}_{r,q}$ contains all values of $\mathbf{p} \in \mathbb{P}$ such that at least q of the functions $s^{\varepsilon}_{i,r}(\mathbf{p})$ are smaller than 0, whereas $\mathbf{\Theta}^{\varepsilon,+}_{r,q}$ contains all values of $\mathbf{p} \in \mathbb{P}$ such that at least q of the functions $s^{\varepsilon}_{i,r}(\mathbf{p})$ are larger than 0.

When the model studied is driven by an input u_t, one may obtain a similar confidence region by substituting $c^{u}_{t-s,s}(\mathbf{p}) = u_{t-s}(\mathbf{p})\,\varepsilon_t(\mathbf{p})$ for $c^{\varepsilon}_{t-r,r}(\mathbf{p}) = \varepsilon_{t-r}(\mathbf{p})\,\varepsilon_t$ in the procedure above, thus replacing autocorrelations by intercorrelations. One then computes a set $\mathbf{\Theta}^{u}_{s,q}$, again such that
$$\Pr\left(\mathbf{p}^* \in \mathbf{\Theta}^{u}_{s,q}\right) = 1 - 2q/m. \tag{11}$$
The fact that the set $\mathbf{\Theta}^{\varepsilon}_{r,q}$ (or $\mathbf{\Theta}^{u}_{r,q}$) obtained by this approach is exact does not mean that its volume is minimal, and the resulting confidence region may turn out to be much too large to be useful. One may then intersect several such regions. For a given value of q and m, assume that n_ε confidence regions $\mathbf{\Theta}^{\varepsilon}_{r,q}$ and n_u confidence regions $\mathbf{\Theta}^{u}_{s,q}$ have been obtained for n_ε values of r and n_u values of s. The probability that $\mathbf{p}^*$ belongs to the intersection $\mathbf{\Theta}$ of these $(n_\varepsilon + n_u)$ regions then satisfies
$$\Pr\left(\mathbf{p}^* \in \mathbf{\Theta}\right) \geqslant 1 - (n_\varepsilon + n_u)2q/m. \tag{12}$$
The price to be paid for taking the intersection of several confidence regions is that the probability that $\mathbf{p}^*$ belongs to the resulting confidence region is no longer known exactly, as only a lower bound for this probability is available.

3. GUARANTEED CHARACTERIZATION VIA INTERVAL ANALYSIS

For the LSCR approach (as well as for the other approaches to build confidence regions), a key issue is the *guaranteed* and *global* numerical characterization of the confidence region. However, and rather surprisingly, this issue seems seldom addressed outside the community of bounded-error estimation. We concentrate here on LSCR confidence regions, but other types of confidence regions could benefit of a similar treatment.

3.1 Problem statement

In LSCR, one has to characterize an intersection of sets defined as
$$\mathbf{\Psi}_q = \left\{\mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q\right\}, \tag{13}$$
where $\tau_i(\mathbf{p})$ is some *indicator* function
$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } f_i(\mathbf{p}) \geqslant 0, \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$
and where $f_i(\mathbf{p})$ depends on the model structure, the measurements, and the parameter vector $\mathbf{p}$.

Characterizing $\mathbf{\Psi}_q$ may be alternatively formulated as a *set-inversion* problem Jaulin and Walter [1993]
$$\mathbf{\Psi}_q = \mathbb{P} \cap \tau^{-1}([q, m]), \tag{15}$$
with
$$\tau(\mathbf{p}) = \sum_{i=1}^{m} \tau_i(\mathbf{p}), \tag{16}$$
which may be efficiently solved via interval analysis Moore [1966], Jaulin et al. [2001]. The next sections briefly recall the basic notions of interval analysis required.

3.2 Inclusion functions and centered forms

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

For arithmetic operations,
$$[x] \circ [y] = \{x \circ y \mid x \in [x], \; y \in [y]\}, \tag{17}$$
where $\circ \in \{+, -, \cdot, /\}$. The set (17) 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 $[x] \subset \mathcal{D}$

$$f([x]) = \{f(x) \mid x \in [x]\} \tag{18}$$

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 (18). Nevertheless, in the general case, obtaining (18) requires the (global) minimization and maximization of f over $[x]$, which is a complicated matter.

Interval analysis makes it possible, through the notion of *inclusion functions*, to obtain guaranteed outer-approximations of $f([x])$. An inclusion function $[f]([x])$ of a function $f(x)$ is such that

$$\forall [x] \subset \mathcal{D},\ f([x]) \subset [f]([x]), \tag{19}$$

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 have been considered. The simplest one is the *natural* inclusion function $[f_{\mathrm{n}}]([x])$, 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. The *centered* form $[f_{\mathrm{c}}]([x])$ takes advantage of the mean-value theorem to get a potentially more accurate approximation. It assumes that f is differentiable, and for any $m \in [x]$, is defined as

$$[f_{\mathrm{c}}]([x]) = f(m) + ([x] - m)[f'][x]), \tag{20}$$

where $[f'][x])$ is an inclusion function of the first derivative of f. Usually, m is taken as the midpoint of $[x]$, $m = \mathrm{mid}([x])$.

Inclusion functions are in general pessimistic, *i.e.*, they provide coarse outer-approximations of the range of function, especially for large intervals and when they are many occurrences of the same real variables. 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, *e.g.*, Jaulin et al. [2001] for more details.

3.3 SIVIA

Consider a set-inversion problem where one has to characterize the set

$$\mathbb{X} = [\mathbf{x}] \cap \mathbf{f}^{-1}(\mathbb{Y}), \tag{21}$$

where $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}^m$, $\mathbb{Y} \subset \mathbb{R}^m$, and $[\mathbf{x}] \subset \mathcal{D}$ is some initial search box for $\mathbb{X}$. The aim of the *Set Inverter Via Interval Analysis* (SIVIA) Jaulin and Walter [1993] 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 $[\mathbf{f}]$ 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 $[\mathbf{f}]([\tilde{\mathbf{x}}]) \subset \mathbb{Y}$ then (19) 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 $[\mathbf{f}]([\tilde{\mathbf{x}}]) \cap \mathbb{Y} = \emptyset$ then (19) 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 ε, $[\tilde{\mathbf{x}}]$ 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 3.1 with SIVIA, an inclusion function for τ is necessary, which will be based on inclusion functions for the τ_i's, and consequently for the f_i's.

3.4 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 entails intractable computational complexity, even for a moderate dimension of $\mathbf{p}$.

Contractors Jaulin et al. [2001] partly address this issue. A contractor $\mathcal{C}_{\mathbf{f},\mathbb{Y}}$ associated with the generic set-inversion problem (21) is a function taking a box $[\mathbf{x}]$ as input and returning a box

$$\mathcal{C}_{\mathbf{f},\mathbb{Y}}([\mathbf{x}]) \subset [\mathbf{x}] \tag{22}$$

such that

$$[\mathbf{x}] \cap \mathbb{X} = \mathcal{C}_{\mathbf{f},\mathbb{Y}}([\mathbf{x}]) \cap \mathbb{X}, \tag{23}$$

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.*

In the specific problems considered here, the role of $\mathbf{x}$ is taken by $\mathbf{p}$. The fact that the function τ introduced in (16) is not differentiable forbids the use of most classic contractors, so a specific contractor is needed. The new contractor proposed is implemented in two steps. It assumes that the functions f_i involved in (14) 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 $f_i(\mathbf{p}) < 0$, $i = 1, \ldots, m$. Second, the union of all non-empty intersections of at least q of these boxes is computed to get a possibly contracted box.

Box contraction using the f_i's The first step uses the centered inclusion function of f_i, which, for some $\mathbf{m} \in [\mathbf{p}]$, may be written as

$$[f_{i,\mathrm{c}}]([\mathbf{p}]) = f_i(\mathbf{m}) + ([\mathbf{p}] - \mathbf{m})^{\mathrm{T}}[\mathbf{g}_i]([\mathbf{p}]) \tag{24}$$

$$= f_i(\mathbf{m}) + \sum_{j=1}^{n_{\mathrm{p}}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]), \tag{25}$$

where $\mathbf{g}_i$ is the gradient of f_i. Using (25), we build a contractor $\mathcal{C}_{f_i,[0,\infty[}$ for the set of all values of $\mathbf{p} \in [\mathbf{p}]$ such that $f_i(\mathbf{p}) \geqslant 0$ as follows.

For the k-th component $[p_k]$ of $[\mathbf{p}]$, when $0 \notin [g_{i,k}]([\mathbf{p}])$, $\mathcal{C}_{f_i,[0,\infty[}$ associates the contracted interval

$$[p'_{i,k}] = [p_k] \cap \left(\left(([f_{i,\mathrm{c}}]([\mathbf{p}]) \cap [0,\infty[) - f_i(\mathbf{m}) - \sum_{j=1, j\neq k}^{n_\mathrm{p}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]) \right) / [g_{i,k}]([\mathbf{p}]) + m_k \right). \tag{26}$$

When $0 \in [g_{i,k}]([\mathbf{p}])$, $\mathcal{C}_{f_i,[0,\infty[}$ leaves $[p_k]$ unchanged, *i.e.*,

$$[p'_{i,k}] = [p_k]. \tag{27}$$

Proposition 1. For all $[\mathbf{p}'_i]$, $i = 1, \ldots, m$, built using (26) and (27), one has

$$[\mathbf{p}'_i] \subset [\mathbf{p}] \tag{28}$$

and

$$[\mathbf{p}'_i] \cap f_i^{-1}([0,\infty[) = [\mathbf{p}] \cap f_i^{-1}([0,\infty[). \tag{29}$$

Proof $[\mathbf{p}'_i] \subset [\mathbf{p}]$ is true by construction. To prove (29), it remains to be proven that $[\mathbf{p}] \cap f_i^{-1}([0,\infty[) \subset [\mathbf{p}'_i] \cap f_i^{-1}([0,\infty[)$. Consider $\mathbf{p}^0 \in [\mathbf{p}] \cap f_i^{-1}([0,\infty[)$. For each index k such that $0 \in [g_{i,k}]([\mathbf{p}])$, $[p_k] = \left[p'_{i,k}\right]$, thus $p_k^0 \in \left[p'_{i,k}\right]$, and only the indices k such that $0 \notin [g_{i,k}]([\mathbf{p}])$ have to be considered. For any $\mathbf{m} \in [\mathbf{p}]$, the mean-value theorem implies that there exists $\boldsymbol{\zeta} \in [\mathbf{p}]$ such that

$$f_i(\mathbf{p}^0) = f_i(\mathbf{m}) + \sum_{j=1}^{n_\mathrm{p}} (p_j^0 - m_j)\, g_{i,j}(\boldsymbol{\zeta}). \tag{30}$$

Since $\mathbf{p}^0 \in f_i^{-1}([0,\infty[)$,

$$f_i(\mathbf{p}^0) \in [0,\infty[. \tag{31}$$

Provided that $g_{i,k}(\boldsymbol{\zeta}) \neq 0$, (30) implies that the k-th component of $\mathbf{p}^0$ satisfies

$$p_k^0 = \left(f_i(\mathbf{p}^0) - f_i(\mathbf{m}) - \sum_{j=1, j\neq k}^{n_\mathrm{p}} (p_j^0 - m_j)\, g_{i,j}(\boldsymbol{\zeta}) \right) / g_{i,k}(\boldsymbol{\zeta}) + m_k. \tag{32}$$

Since $[f_{i,\mathrm{c}}](\cdot)$, introduced in (25), is an inclusion function for $f_i(\cdot)$, (31) combined with the fact that $\mathbf{p}^0 \in [\mathbf{p}]$ and $\boldsymbol{\zeta} \in [\mathbf{p}]$, implies that

$$p_k^0 \in [p_k] \cap \left(\left((([f_{i,\mathrm{c}}]([\mathbf{p}]) \cap [0,\infty[) - f_i(\mathbf{m}) - \sum_{j=1, j\neq k}^{n_\mathrm{p}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]) \right) / [g_{i,k}]([\mathbf{p}]) + m_k \right) \tag{33}$$

$$\in [p'_k], \tag{34}$$

provided that $0 \notin [g_{i,k}]([\mathbf{p}])$. Thus for each k, $p_k^0 \in [p'_k]$. Since $\mathbf{p}^0$ also belongs to $f_i^{-1}([0,\infty[)$, $\mathbf{p}^0 \in [\mathbf{p}'_i] \cap f_i^{-1}([0,\infty[)$, which completes the proof. ◇

Considering the m functions f_i and applying all the contractors $\mathcal{C}_{f_i,[0,\infty[}$, $i = 1, \ldots, n$, to $[\mathbf{p}]$, one obtains a list of m possibly contracted boxes

```
1  [p] = ∅;
2  Reindex the boxes [p_i] in such a way that
     p̲_1 ⩽ p̲_2 ⩽ ··· ⩽ p̲_n;
3  For i = q to n
4      if Σ_{j=1}^{n} (p̲_i ∈ [p_j]) ⩾ q
5          p̲ = p̲_i; break;
6  Reindex the boxes [p_i] in such a way that
     p̄_1 ⩾ p̄_2 ⩾ ··· ⩾ p̄_n
7  For i = q to n
8      if Σ_{j=1}^{n} (p̄_i ∈ [p_j]) ⩾ q
9          p̄ = p̄_i; break;
```

Algorithm 1. $[p] = q$-relaxed intersection $([p_1], \ldots, [p_n])$

$$\mathcal{L} = \left\{ \mathcal{C}_{f_1,[0,\infty[}([\mathbf{p}]), \ldots, \mathcal{C}_{f_m,[0,\infty[}([\mathbf{p}]) \right\} \tag{35}$$

$$= \{[\mathbf{p}'_1], \ldots, [\mathbf{p}'_m]\}. \tag{36}$$

Some of them may be empty, in which case, $[\mathbf{p}'_i] = \emptyset$ indicates that there is no $\mathbf{p} \in [\mathbf{p}]$ such that $f_i(\mathbf{p}) \geqslant 0$. Our aim is to evaluate a subbox $[\mathbf{p}']$ of $[\mathbf{p}]$ such that $\boldsymbol{\Psi}_q \cap [\mathbf{p}'] = \boldsymbol{\Psi}_q \cap [\mathbf{p}]$.

Building a q-relaxed intersection During the second step, the contractor builds a box $[\mathbf{p}']$ enclosing the q-relaxed intersection $\mathcal{P}$ Jaulin and Walter [2002], Jaulin [2009, 2011] of the boxes in $\mathcal{L} = \{[\mathbf{p}'_1], \ldots, [\mathbf{p}'_m]\}$, *i.e.*, the union of all intersections of at least q boxes in $\mathcal{L}$

$$\mathcal{P} = \bigcap_{j \in \{1,\ldots,m\}}^{q} [\mathbf{p}'_j]. \tag{37}$$

$$= \bigcup_{\substack{J \subset [1,\ldots,m] \\ \mathrm{card}(J) \geqslant q}} \bigcap_{j \in J} [\mathbf{p}'_j], \tag{38}$$

and satisfying

$$\mathcal{P} \subset [\mathbf{p}'] \subset [\mathbf{p}]. \tag{39}$$

Proposition 2. For any box $[\mathbf{p}']$, satisfying (39), one has

$$\boldsymbol{\Psi}_q \cap [\mathbf{p}'] = \boldsymbol{\Psi}_q \cap [\mathbf{p}],$$

with $\boldsymbol{\Psi}_q$ as defined in (13).

Proof Assume that there exists $\mathbf{p}_0 \in [\mathbf{p}]$ such that $\mathbf{p}_0 \in \boldsymbol{\Psi}_q \cap [\mathbf{p}]$ but $\mathbf{p}_0 \notin \boldsymbol{\Psi}_q \cap [\mathbf{p}']$. Since $\mathbf{p}_0 \in \boldsymbol{\Psi}_q \cap [\mathbf{p}]$, $\mathbf{p}_0 \in \boldsymbol{\Psi}_q$. According to (13), $\sum_{i=1}^{m} \tau_i(\mathbf{p}_0) \geqslant q$. There are thus at least q functions τ_i such that $\tau_i(\mathbf{p}_0) \geqslant 1$. Assume, without loss of generality, that $\tau_1(\mathbf{p}_0) \geqslant 1, \ldots, \tau_q(\mathbf{p}_0) \geqslant 1$. Since $\tau_i(\mathbf{p}_0) \geqslant 1$, $i = 1, \ldots, q$, by definition of $\mathcal{C}_{f_i,[0,\infty[}$, one has $\mathbf{p}_0 \in [\mathbf{p}'_i]$, $i = 1, \ldots, q$ and $\mathbf{p}_0 \in \bigcap_{i=1,\ldots,q} [\mathbf{p}'_i]$. By definition of $\mathcal{P}$ and $[\mathbf{p}']$, $\mathbf{p}_0 \in \bigcap_{i=1,\ldots,q} [\mathbf{p}'_i] \subset \mathcal{P} \subset [\mathbf{p}']$, which contradicts the initial assumption. ◇

Algorithm 1 formalizes a computation carried out on an example in Jaulin [2011]. 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 $\mathcal{L} = \{[p_1], \ldots, [p_m]\}$ of m 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 (37) 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.

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

4. EXAMPLES

Two examples are considered in what follows to illustrate the guaranteed characterization of exact confidence regions provided by LSCR.

All computations were carried out with Intlab Rump [2001], the interval-analysis toolbox for Matlab.

4.1 Exponential model

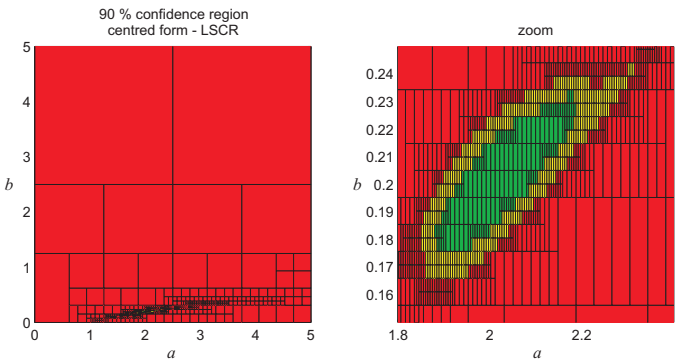

Fig. 1. Paving of search space obtained in the exponential case for the characterization of $\mathbf{\Theta}^{\varepsilon}_{r=1,q=3}$

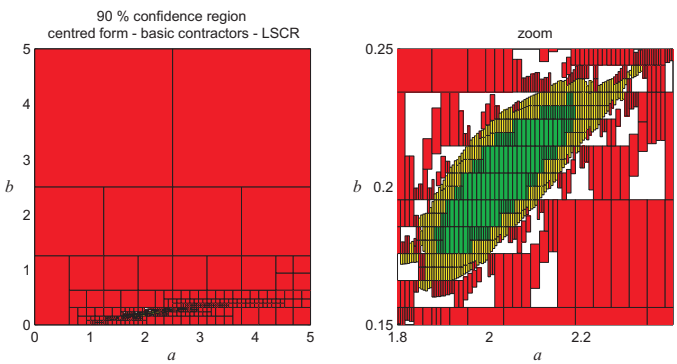

Fig. 2. Subpavings of search space obtained in the exponential case for the characterization of $\mathbf{\Theta}^{\varepsilon}_{r=1,q=3}$, with the contractors of Section 3.4

Consider first a system such that

$$y_t = a^* \exp\left(-b^* t\right) + w_t \tag{40}$$

where, for $t = 0, T, \ldots, (n-1)\,T$, the w_t's are assumed to be realizations of iid $\mathcal{N}\left(0, \sigma^2\right)$ variables.

Take as a model

$$y_t^{\mathrm{m}}\left(\mathbf{p}\right) = a \exp\left(-bt\right), \tag{41}$$

which is nonlinear in $\mathbf{p} = (a, b)^{\mathrm{T}}$. For the true value of the parameter vector, the prediction errors defined as

$$\varepsilon_t\left(\mathbf{p}\right) = y_t - y_t^{\mathrm{m}}\left(\mathbf{p}\right), \tag{42}$$

with $y_t^{\mathrm{m}}\left(\mathbf{p}\right)$ taken from (41), are iid $\mathcal{N}\left(0, \sigma^2\right)$. Their distribution is thus symmetric, and LSCR applies.

The characterization of $\mathbf{\Theta}^{\varepsilon}_{r,q}$ (or $\mathbf{\Theta}^{u}_{s,q}$) is addressed using SIVIA. For that purpose, inclusion functions for the indicator functions $\tau_i^{\varepsilon,-}$ and $\tau_i^{\varepsilon,+}$ introduced in (9) and (10) are provided

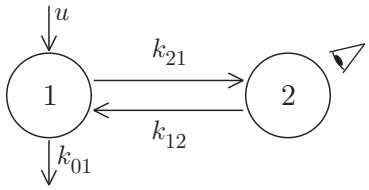

Fig. 3. Two-compartment model

$$\left[\tau_i^{\varepsilon,-}\right]\left([\mathbf{p}]\right) = \begin{cases} 1 & \text{if } \sup\left(\left[s_{i,r}^{\varepsilon}\right]\left([\mathbf{p}]\right)\right) < 0, \\ 0 & \text{if } \inf\left(\left[s_{i,r}^{\varepsilon}\right]\left([\mathbf{p}]\right)\right) \geqslant 0, \\ [0,1] & \text{else,} \end{cases} \tag{43}$$

and

$$\left[\tau_i^{\varepsilon,+}\right]\left([\mathbf{p}]\right) = \begin{cases} 1 & \text{if } \inf\left(\left[s_{i,r}^{\varepsilon}\right]\left([\mathbf{p}]\right)\right) \geqslant 0, \\ 0 & \text{if } \sup\left(\left[s_{i,r}^{\varepsilon}\right]\left([\mathbf{p}]\right)\right) < 0, \\ [0,1] & \text{else.} \end{cases} \tag{44}$$

In (43) and (44), $\left[s_{i,r}^{\varepsilon}\right]$ is an inclusion function for $s_{i,r}^{\varepsilon}$.

Consider $n = 64$ measurements collected regularly with a period $T = 0.1$ s, with $a^* = 2$, $b^* = 0.2$, and $\sigma^2 = 0.01$. The initial search set in parameter space is taken as $\mathbb{P} = [0,5] \times [0,5]$, and the precision parameter as $\varepsilon = 0.005$. Figure 1 represents guaranteed inner and outer approximations of the set $\mathbf{\Theta}^{\varepsilon}_{r=1,q=3}$ obtained for $q = 3$, which corresponds to a 90% confidence region. This set turns out to contain the true value of the parameter vector, which should of course not always be true.

With the same experimental conditions, Figure 1 represents guaranteed inner and outer approximations of the set $\mathbf{\Theta}^{\varepsilon}_{r=1,q=3}$ obtained for $q = 3$, when the contractors of Section 3.4 are employed. As can be seen, the contractors are only efficient for small boxes close to the solution set.

4.2 Compartmental model

Consider now the two-compartment model described by Figure 3. The system output satisfies

$$y_t = \alpha\left(\mathbf{p}^*\right)\left(\exp\left(\lambda_1\left(\mathbf{p}^*\right)t\right) - \exp\left(\lambda_2\left(\mathbf{p}^*\right)t\right)\right) + w_t, \tag{45}$$

where $\mathbf{p} = \left(k_{01}, k_{12}, k_{21}\right)^{\mathrm{T}}$,

$$\alpha\left(\mathbf{p}\right) = k_{21} / \sqrt{\left(k_{01} - k_{12} + k_{21}\right)^2 + 4k_{12}k_{21}}, \tag{46}$$

$$\lambda_{1,2}\left(\mathbf{p}\right) = -\frac{1}{2}\left(\left(k_{01} + k_{12} + k_{21}\right) \pm \left(\left(k_{01} - k_{12} + k_{21}\right)^2 + 4k_{12}k_{21}\right)^{-1/2}\right) \tag{47}$$

and the w_t's are realizations of iid $\mathcal{N}\left(0, \sigma^2\right)$ variables, for $t = 0, T, \ldots, (n-1)\,T$. Data have been generated with $\mathbf{p}^* = \left(1, 0.25, 0.5\right)^{\mathrm{T}}$. The variance of the measurement noise is $\sigma^2 = 10^{-4}$. The sampling period is $T = 0.2$ s, and $n = 64$. To facilitate illustration, only k_{01} et k_{12} are estimated. The value k_{21}^* of k_{21} is assumed known.

Since the measurement noise is additive, the LSCR method applies directly, as in the previous example. For a given value $\mathbf{p} = \left(k_{01}, k_{12}, k_{21}^*\right)^{\mathrm{T}}$ of the parameter vector, the prediction errors are $\varepsilon_t\left(\mathbf{p}\right) = y_t - y_t^{\mathrm{m}}\left(\mathbf{p}\right)$, with

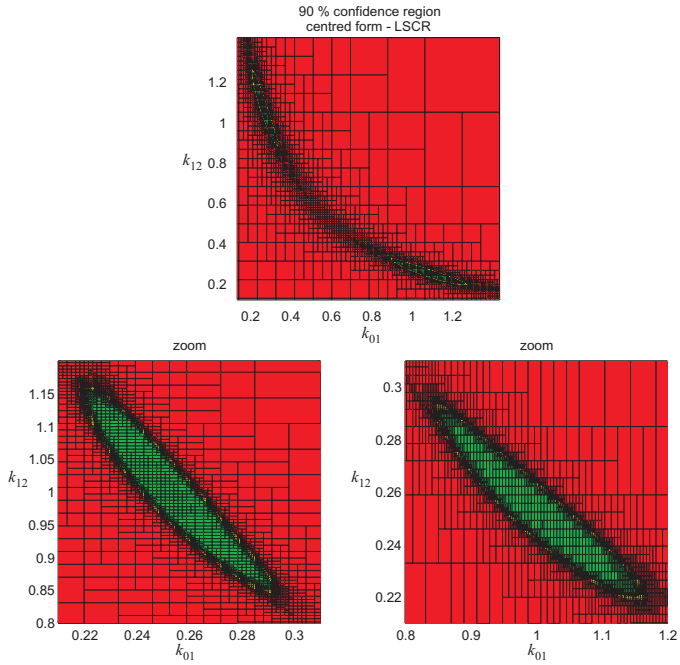

Fig. 4. Paving of search space $\mathbb{P} = [0,5] \times [0,5]$ obtained in the two-compartment model case for the characterization of the confidence region $\mathbf{\Theta}^{\varepsilon}_{r,q}$; the two bottom subfigures are zooms on the two disconnected components of $\mathbf{\Theta}^{\varepsilon}_{r,q}$

$y^{\mathrm{m}}_t(\mathbf{p}) = \alpha(\mathbf{p})(\exp(\lambda_1(\mathbf{p})t) - \exp(\lambda_2(\mathbf{p})t))$, for $t = 0, T, \ldots, (n-1)T$.

Here, the set $\mathbf{\Theta}^{\varepsilon}_{r,q}$ has been characterized with $r = 1$ and $q = 3$, which corresponds to a 90 % confidence region, see Figure 4. The initial search set in parameter space is $\mathbb{P} = [0,5] \times [0,5]$ and $\varepsilon = 0.001$. The results show that the confidence region consists of two disconnected subsets, a consequence of the lack of global identifiability of the considered model, which is only locally identifiable. The values of k_{01} and k_{12} may be exchanged without changing the model output. Figure 4 (bottom) focuses on the two confidence subsets, one of which turns out to contain the actual value of the two unknown parameters, although this is not guaranteed, of course.

5. CONCLUSIONS AND PERSPECTIVES

Interval analysis provides tools to evaluate guaranteed inner and outer-approximations of non-asymptotic confidence regions defined by LSCR. Illustrations have been provided for two models with outputs nonlinear in their parameters. Applications to the characterization of confidence regions defined by SPS may be found in Kieffer and Walter [2012].

Accurate inclusion functions are particularly difficult to obtain for the functions involved in LSCR, due to the many occurrences of the parameters involved in the evaluation of (4). Symbolic manipulations of the expressions involved to reduce the number of occurrences of the parameters may be particularly useful to improve the efficiency of SIVIA and to design better contractors than those considered in Section 3.4.1.

The availability of efficient contractors is mandatory to address the problems of characterizing confidence regions for models with a large number of parameters.

REFERENCES

M. C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005.

B. C. Csáji, M. C. Campi, and E. Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proc. IFAC Symposium on System Identification*, pages 227–232, Brussels, Belgium, 2012.

M. Dalai, E. Weyer, and M. C. Campi. Parameter identification for nonlinear systems: Guaranteed confidence regions through LSCR. *Automatica*, 43:1418 – 1425, 2007.

L. Gordon. Completely separating groups in subsampling. *Annals of Statistics*, 2(3):572–578, 1974.

L. Jaulin. Robust set membership state estimation; application to underwater robotics. *Automatica*, 45(1):202–206, 2009.

L. Jaulin. Set-membership localization with probabilistic errors. *Robotics and Autonomous Systems*, 59(6):489–495, 2011.

L. Jaulin and E. Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29 (4):1053–1064, 1993.

L. Jaulin and E. Walter. Guaranteed robust nonlinear minimax estimation. *IEEE Transaction on Automatic Control*, 47(11):1857–1864, 2002.

L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis*. Springer-Verlag, London, 2001.

M. Kieffer and É. Walter. Guaranteed characterization of exact non-asymptotic confidence regions. *Automatica*, 2012. submitted.

R. E. Moore. *Interval Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

S. M. Rump. INTLAB - INTerval LABoratory. In J. Grabmeier, E. Kaltofen, and V. Weispfennig, editors, *Handbook of Computer Algebra: Foundations, Applications, Systems*. Springer-Verlag, Heidelberg, Germany, 2001.