Guaranteed characterization
of exact non-asymptotic confidence regions
as defined by LSCR and SPS

Michel Kieffer [a,b], Eric Walter [a],

[a]*L2S, CNRS, Supelec, Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France*

[b]*Institut Universitaire de France, 103 bld Saint-Michel, 75005 Paris*

Abstract

In parameter estimation, it is often desirable to supplement the estimates with an assessment of their quality. A new family of methods proposed by Campi *et al.* for this purpose is particularly attractive, as it makes it possible to obtain exact, non-asymptotic confidence regions under relatively mild assumptions on the noise distribution. A bottleneck of this approach, however, is the numerical characterization of these confidence regions. So far, it has been carried out by gridding, which provides no guarantee as to its results and is only applicable to low dimensional spaces. The aim of this paper is to show how interval analysis can contribute to removing this bottleneck.

Key words: Confidence regions; Interval analysis; Nonlinear system identification.

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

$$\widehat{\mathbf{p}} = \arg\min_{\mathbf{p}} J(\mathbf{p}). \qquad (1)$$

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

Even if a single numerical vector $\widehat{\mathbf{p}}$ is obtained and if $\mathbf{y}$ and $\mathbf{y}^{\mathrm{m}}(\widehat{\mathbf{p}})$ are reassuringly similar, it would be naive to consider $\widehat{\mathbf{p}}$ as the final answer to the estimation problem. One should instead attempt to attach some quality tag to $\widehat{\mathbf{p}}$ 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. A key issue is drawing conclusions that are as little prejudiced as possible, and the approach recently proposed by Campi *et al.* for this purpose [1–3] is particularly attractive, as it makes it possible to obtain exact, non-asymptotic confidence regions under relatively mild assumptions on the noise distribution. A difficulty with this approach, however, is the numerical characterization of these confidence regions. So far, it has been carried out by gridding, which provides no guarantee as to its results and is only applicable to low dimensional spaces.

The aim of this paper is to show how interval analysis can contribute to adressing this difficulty, by providing guaranteed results, as well as results in high dimensional spaces. The approaches *Leave-out Sign-dominant Correlated Regions* (LSCR) [1,3] and *Sign-Perturbed Sums*

(SPS) [2] recently proposed by Campi *et al.* are briefly recalled in Section 2. Section 3 shows how interval analysis can be used to characterize the exact confidence regions defined by LSCR and SPS in a global and guaranteed way. Examples are treated in Section 4, and conclusions are drawn in Section 5.

2 LSCR and SPS

The most striking feature of LSCR and SPS is that they avoid a large number of the usual assumptions about the noise corrupting the data yet provide an *exact* characterization of parameter uncertainty in *non-asymptotic* conditions. It is not 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. LSCR and SPS are summarized in the next two sections. Both require the noise samples to be independently distributed with distributions symmetric with respect to zero.

2.1 LSCR

LSCR [1] defines a region $\boldsymbol{\Theta}$ to which the parameter vector $\mathbf{p}^*$ of the true system belongs with a 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. It may, for instance, be the difference between some output measurement y_t, and the corresponding model output $y_t^{\mathrm{m}}(\mathbf{p})$. 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

$$c_{t-r,r}^{\varepsilon}(\mathbf{p}) = \varepsilon_{t-r}(\mathbf{p})\, \varepsilon_t(\mathbf{p}), \qquad (2)$$

(3) Compute

$$s_{i,r}^{\varepsilon}(\mathbf{p}) = \sum_{k \in \mathbb{I}_i} c_{k,r}^{\varepsilon}(\mathbf{p}), \; i = 1, \ldots, m, \qquad (3)$$

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 $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ such that *at least* q of the functions $s_{i,r}^{\varepsilon}(\mathbf{p})$ are *larger* than 0 and *at least* q are *smaller* than 0.

The probability that $\mathbf{p}^*$ belongs to $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ is

$$\Pr\left(\mathbf{p}^* \in \boldsymbol{\Theta}_{r,q}^{\varepsilon}\right) = 1 - 2q/m. \qquad (4)$$

The shape and size of $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ 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 [4].

The set $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ may be defined more formally as

$$\boldsymbol{\Theta}_{r,q}^{\varepsilon} = \boldsymbol{\Theta}_{r,q}^{\varepsilon,1} \cap \boldsymbol{\Theta}_{r,q}^{\varepsilon,2}, \qquad (5)$$

with, for $j = 1, 2$,

$$\boldsymbol{\Theta}_{r,q}^{\varepsilon,j} = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i^{\varepsilon,j}(\mathbf{p}) \geqslant q \right\}, \qquad (6)$$

where $\mathbb{P}$ is the prior domain for $\mathbf{p}$ and where

$$\tau_i^{\varepsilon,j}(\mathbf{p}) = \begin{cases} 1 & \text{if } (-1)^j s_{i,r}^{\varepsilon}(\mathbf{p}) \geqslant 0, \\ 0 & \text{else.} \end{cases} \qquad (7)$$

The set $\boldsymbol{\Theta}_{r,q}^{\varepsilon,1}$ contains all values of $\mathbf{p} \in \mathbb{P}$ such that at least q of the functions $s_{i,r}^{\varepsilon}(\mathbf{p})$ are smaller than 0, whereas $\boldsymbol{\Theta}_{r,q}^{\varepsilon,2}$ contains all values of $\mathbf{p} \in \mathbb{P}$ such that at least q of the functions $s_{i,r}^{\varepsilon}(\mathbf{p})$ are larger than 0.

When the model studied is driven by an input u_t, one may obtain a similar confidence region by substituting $c_{t-s,s}^u(\mathbf{p}) = u_{t-s}(\mathbf{p})\, \varepsilon_t(\mathbf{p})$ for $c_{t-r,r}^{\varepsilon}(\mathbf{p}) = \varepsilon_{t-r}(\mathbf{p})\, \varepsilon_t$ in the procedure above, thus replacing autocorrelations by intercorrelations. One then computes a set $\boldsymbol{\Theta}_{s,q}^u$, again such that $\Pr\left(\mathbf{p}^* \in \boldsymbol{\Theta}_{s,q}^u\right) = 1 - 2q/m$.

The fact that the set $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ (or $\boldsymbol{\Theta}_{r,q}^u$) 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 $\boldsymbol{\Theta}_{r,q}^{\varepsilon}$ and n_u confidence regions $\boldsymbol{\Theta}_{s,q}^u$ have been obtained for n_ε values of r and n_u values of s. The probability that $\mathbf{p}^*$ belongs to the intersection $\boldsymbol{\Theta}$ of these $(n_\varepsilon + n_u)$ regions then satisfies $\Pr\left(\mathbf{p}^* \in \boldsymbol{\Theta}\right) \geqslant 1 - (n_\varepsilon + n_u)2q/m$. 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.

2.2 SPS

SPS [2] also 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

$$y_t = \boldsymbol{\varphi}_t^{\mathrm{T}} \mathbf{p}^* + w_t, \, t = 1, \ldots, n, \qquad (8)$$

with $\boldsymbol{\varphi}_t$ a known regression vector that does not depend on the unknown parameters. It computes an exact

confidence region for $\mathbf{p}^*$ around the least-squares estimate $\widehat{\mathbf{p}}$, which is the solution to the *normal equations* $\sum_{t=1}^{n} \boldsymbol{\varphi}_t \left(y_t - \boldsymbol{\varphi}_t^{\mathrm{T}} \widehat{\mathbf{p}} \right) = \mathbf{0}$. For a generic $\mathbf{p}$, define

$$\mathbf{s}_0 \left(\mathbf{p} \right) = \sum_{t=1}^{n} \boldsymbol{\varphi}_t \left(y_t - \boldsymbol{\varphi}_t^{\mathrm{T}} \mathbf{p} \right), \tag{9}$$

and the *sign-perturbed sums*

$$\mathbf{s}_i \left(\mathbf{p} \right) = \sum_{t=1}^{n} \alpha_{i,t} \boldsymbol{\varphi}_t \left(y_t - \boldsymbol{\varphi}_t^{\mathrm{T}} \mathbf{p} \right), \tag{10}$$

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

$$z_i \left(\mathbf{p} \right) = \| \mathbf{s}_i \left(\mathbf{p} \right) \|_2^2, i = 0, \ldots, m-1. \tag{11}$$

A confidence region $\boldsymbol{\Sigma}_q$ 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 [2], it has been shown that $\mathbf{p}^*$ belongs to $\boldsymbol{\Sigma}_q$ with *exact* probability $1 - q/m$. $\boldsymbol{\Sigma}_q$ may be defined more formally as

$$\boldsymbol{\Sigma}_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m-1} \tau_i \left(\mathbf{p} \right) \geqslant q \right\} \tag{12}$$

where

$$\tau_i \left(\mathbf{p} \right) = \begin{cases} 1 & \text{if } z_i \left(\mathbf{p} \right) - z_0 \left(\mathbf{p} \right) > 0, \\ 0 & \text{else.} \end{cases} \tag{13}$$

This is justified by the fact that if $\sum_{i=1}^{m-1} \tau_i \left(\mathbf{p} \right) \geqslant q$, then one has $\tau_i \left(\mathbf{p} \right) = 1$ 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 $\left(z_i \left(\mathbf{p} \right) \right)_{i=0}^{m-1}$.

3 Guaranteed characterization via interval analysis

In LSCR and SPS, one has to characterize a set or intersections of sets defined as

$$\boldsymbol{\Psi}_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i \left(\mathbf{p} \right) \geqslant q \right\}, \tag{14}$$

where $\tau_i \left(\mathbf{p} \right)$ is some *indicator* function

$$\tau_i \left(\mathbf{p} \right) = \begin{cases} 1 & \text{if } f_i \left(\mathbf{p} \right) \geqslant 0, \\ 0 & \text{else,} \end{cases} \tag{15}$$

and where $f_i \left(\mathbf{p} \right)$ depends on the model structure, the measurements, and the parameter vector $\mathbf{p}$.

Characterizing $\boldsymbol{\Psi}_q$ may be alternatively formulated as a *set-inversion* [8] problem

$$\boldsymbol{\Psi}_q = \mathbb{P} \cap \tau^{-1} \left([q, m] \right), \tag{16}$$

with

$$\tau \left(\mathbf{p} \right) = \sum_{i=1}^{m} \tau_i \left(\mathbf{p} \right), \tag{17}$$

which may be efficiently solved via interval analysis [7,9] using the SIVIA algorithm [7]. For that purpose, inclusion functions for the τ_i's and consequently for the f_i's are required.

SIVIA recursively partitions $\mathbb{P}$ into boxes (vectors of intervals) proved to belong to $\boldsymbol{\Psi}_q$, boxes proved to have no intersection with $\boldsymbol{\Psi}_q$, and *undetermined* boxes for which no conclusion can be obtained. SIVIA bisects undetermined boxes until their width is less than some precision parameter ε.

3.1 Contractors for guaranteed characterization

Indetermination often results from range overestimation by inclusion functions. As a consequence, boxes have to be bisected by SIVIA many times to allow one to conclude on the position of the resulting boxes with respect to $\boldsymbol{\Psi}_q$. This may entail an intractable computational complexity, even for a moderate dimension of $\mathbf{p}$.

Contractors [7] partly address this issue. Consider a set-inversion problem where one has to characterize the set

$$\mathbb{X} = [\mathbf{x}] \cap \mathbf{f}^{-1} \left(\mathbb{Y} \right), \tag{18}$$

with $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}^m$, $\mathbb{Y} \subset \mathbb{R}^m$, and $[\mathbf{x}] \subset \mathcal{D}$ some initial search box for $\mathbb{X}$. A contractor $\mathcal{C}_{\mathbf{f}, \mathbb{Y}}$ associated with the generic set-inversion problem (18) is a function taking a box $[\mathbf{x}]$ as input and returning a box

$$\mathcal{C}_{\mathbf{f}, \mathbb{Y}} \left([\mathbf{x}] \right) \subset [\mathbf{x}] \tag{19}$$

such that

$$[\mathbf{x}] \cap \mathbb{X} = \mathcal{C}_{\mathbf{f}, \mathbb{Y}} \left([\mathbf{x}] \right) \cap \mathbb{X}, \tag{20}$$

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 bisection. Various 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.* [7].

In the problems considered here, the role of $\mathbf{x}$ is taken by $\mathbf{p}$. The fact that the function τ introduced in (17) is not differentiable forbids the use of most classic contractors, so a specific contractor is needed. The new contractor proposed here is implemented in two steps. It assumes that the functions f_i involved in (15) are differentiable.

3

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 $f_i(\mathbf{p}) < 0$, $i = 1, \ldots, m$, see Sections 3.1.1 and 3.1.2. 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 3.1.3.

3.1.1 Contractor for LSCR and SPS

The first step uses the *centered form* of f_i, which, for some $\mathbf{m} \in [\mathbf{p}]$, may be written as

$$[f_{i,\mathrm{c}}]([\mathbf{p}]) = f_i(\mathbf{m}) + ([\mathbf{p}] - \mathbf{m})^\mathrm{T}[\mathbf{g}_i]([\mathbf{p}]) \qquad (21)$$

$$= f_i(\mathbf{m}) + \sum_{j=1}^{n_\mathrm{p}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]), \quad (22)$$

where $\mathbf{g}_i$ is the gradient of f_i and $[\mathbf{g}_i]([\mathbf{p}])$ is the natural inclusion function for $\mathbf{g}_i$, see [7]. Using (22), 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.

With 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

$$\begin{aligned}
[p'_{i,k}] = [p_k] \cap \Bigg(\Bigg(\Big(([f_{i,\mathrm{c}}]([\mathbf{p}]) \cap [0,\infty[) - f_i(\mathbf{m}) \\
- \sum_{j=1, j \neq k}^{n_\mathrm{p}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]) \Big) / [g_{i,k}]([\mathbf{p}]) + m_k \Bigg).
\end{aligned}$$
$$(23)$$

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

$$[p'_{i,k}] = [p_k]. \qquad (24)$$

Due to the pessimism of centered forms on large boxes, the contractor $\mathcal{C}_{f_i,[0,\infty[}$ becomes efficient only when the box to contract is small enough.

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

$$\mathcal{L} = \{\mathcal{C}_{f_1,[0,\infty[}([\mathbf{p}]), \ldots, \mathcal{C}_{f_m,[0,\infty[}([\mathbf{p}])\} \qquad (25)$$
$$= \{[\mathbf{p}'_1], \ldots, [\mathbf{p}'_m]\}. \qquad (26)$$

Some of these boxes 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 $\mathbf{\Psi}_q \cap [\mathbf{p}'] = \mathbf{\Psi}_q \cap [\mathbf{p}]$.

3.1.2 Contractor for SPS

We take advantage of the fact that the functions $\mathbf{s}_i(\mathbf{p})$, $i = 0, \ldots, m$ are affine in $\mathbf{p}$ to reduce the number of occurrences of $\mathbf{p}$ in their formal expression, and thus to reduce the pessimism of the corresponding inclusion functions. Equation (9) is rewritten as

$$\mathbf{s}_0(\mathbf{p}) = \mathbf{b}_0 - \mathbf{A}_0\mathbf{p} \qquad (27)$$

with $\mathbf{b}_0 = \sum_{t=1}^n y_t\boldsymbol{\varphi}_t$ and $\mathbf{A}_0 = \sum_{t=1}^n \boldsymbol{\varphi}_t\boldsymbol{\varphi}_t^\mathrm{T}$. Similarly, (10) is rewritten as

$$\mathbf{s}_i(\mathbf{p}) = \mathbf{b}_i - \mathbf{A}_i\mathbf{p} \qquad (28)$$

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

Equations (11), (27), (28), and the fact that the $\mathbf{A}_i$'s are symmetric imply that

$$\begin{aligned}
z_i(\mathbf{p}) - z_0(\mathbf{p}) = \mathbf{p}^\mathrm{T}\left(\mathbf{A}_i^2 - \mathbf{A}_0^2\right)\mathbf{p} - 2\left(\mathbf{b}_i^\mathrm{T}\mathbf{A}_i - \mathbf{b}_0^\mathrm{T}\mathbf{A}_0\right)\mathbf{p} \\
+ \left(\mathbf{b}_i^\mathrm{T}\mathbf{b}_i - \mathbf{b}_0^\mathrm{T}\mathbf{b}_0\right).
\end{aligned}$$
$$(29)$$

The matrices $\mathbf{A}_i^2 - \mathbf{A}_0^2$ are symmetric and may thus be diagonalized as $\mathbf{A}_i^2 - \mathbf{A}_0^2 = \mathbf{U}_i^\mathrm{T}\mathbf{D}_i\mathbf{U}_i$, where $\mathbf{U}_i$ is an orthonormal matrix (*i.e.*, such that $\mathbf{U}_i^\mathrm{T} = \mathbf{U}_i^{-1}$), and $\mathbf{D}_i = \mathrm{diag}\left(d_{i,1}, \ldots, d_{i,n_\mathrm{p}}\right)$ is a diagonal matrix. With the change of variables $\boldsymbol{\pi} = \mathbf{U}_i\mathbf{p}$, (29) becomes

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \boldsymbol{\pi}^\mathrm{T}\mathbf{D}_i\boldsymbol{\pi} - 2\boldsymbol{\beta}_i^\mathrm{T}\boldsymbol{\pi} + \gamma_i, \qquad (30)$$

where $\boldsymbol{\beta}_i^\mathrm{T} = \left(\mathbf{b}_i^\mathrm{T}\mathbf{A}_i - \mathbf{b}_0^\mathrm{T}\mathbf{A}_0\right)\mathbf{U}_i^\mathrm{T}$ and $\gamma_i = \mathbf{b}_i^\mathrm{T}\mathbf{b}_i - \mathbf{b}_0^\mathrm{T}\mathbf{b}_0$. Then, provided that $d_{i,j} \neq 0$ for $j = 1, \ldots, n_\mathrm{p}$, (30) can be rewritten as

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \sum_{j=1}^{n_\mathrm{p}} d_{i,j}\left(\pi_j - \frac{\beta_{i,j}}{d_{i,j}}\right)^2 + \gamma_i - \sum_{j=1}^{n_\mathrm{p}} \frac{\beta_{i,j}^2}{d_{i,j}}.$$
$$(31)$$

Let $[\boldsymbol{\pi}] = \mathbf{U}_i[\mathbf{p}]$. A contractor for $[\pi_j]$ is obtained from (31) as follows

$$\begin{aligned}
[\pi'_j] = [\pi_j] \cap \Bigg\{ \pm\Bigg(\frac{1}{d_{i,j}}\Bigg(\Big(([z_i]([\mathbf{p}]) - [z_0]([\mathbf{p}])) \cap [0,\infty[\Big) \\
- \sum_{\substack{k=1 \\ k \neq j}}^{n_\mathrm{p}} d_{i,k}\left([\pi_k] - \frac{\beta_{i,k}}{d_{i,k}}\right)^2 - \gamma_j + \sum_{k=1}^{n_\mathrm{p}} \frac{\beta_{i,k}^2}{d_{i,k}}\Bigg)\Bigg)^{\frac{1}{2}} + \frac{\beta_{i,j}}{d_{i,j}} \Bigg\}.
\end{aligned}$$
$$(32)$$

From (32), the contractor $\mathcal{C}_{z_i-z_0,[0,\infty[}$ for $[\mathbf{p}]$ is such that

$$[\mathbf{p}'_i] = \mathcal{C}_{z_i-z_0,[0,\infty[}([\mathbf{p}]) = [\mathbf{p}] \cap \left(\mathbf{U}_i^\mathrm{T}[\boldsymbol{\pi}']\right). \qquad (33)$$

When n is large enough and provided that the $\boldsymbol{\varphi}_t$'s have been well designed, it is very unlikely that $\mathbf{A}_i^2 - \mathbf{A}_0^2$ is rank deficient. Should this occur, (31) and (32) would have to be rewritten distinguishing the zero and nonzero $d_{i,j}$'s.

Proposition 1 *Provided that $d_{i,j} \neq 0$ for $j = 1, \ldots, n_p$, for all $[\mathbf{p}'_i]$, $i = 1, \ldots, m - 1$, built using (32) and (33), $[\mathbf{p}'_i] \subset [\mathbf{p}]$ and*

$$[\mathbf{p}'_i] \cap (z_i - z_0)^{-1}([0, \infty[) = [\mathbf{p}] \cap (z_i - z_0)^{-1}([0, \infty[). \tag{34}$$

Proof $[\mathbf{p}'_i] \subset [\mathbf{p}]$ is true by construction. To prove (34), 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 $\boldsymbol{\pi}^0 \in [\boldsymbol{\pi}']$. By definition of $\mathbf{p}^0$, one has

$$\sum_{j=1}^{n_p} d_{i,j} \left(\pi_j^0 - \frac{\beta_{i,j}}{d_{i,j}} \right)^2 + \gamma_i - \sum_{j=1}^{n_p} \frac{\beta_{i,j}^2}{d_{i,j}} = z_i\left(\mathbf{p}^0\right) - z_0\left(\mathbf{p}^0\right)$$
$$\geqslant 0. \tag{35}$$

Since $\boldsymbol{\pi}^0 \in [\boldsymbol{\pi}]$, after some manipulations of (35), one gets for $j = 1, \ldots, n_p$

$$\pi_j^0 \in [\pi_j] \cap \left\{ \pm \left(\frac{1}{d_{i,j}} \left(\left(\left([z_i]\left([\mathbf{p}]\right) - [z_0]\left([\mathbf{p}]\right) \right) \cap [0, \infty[\right) \right. \right. \right.$$
$$\left. \left. - \sum_{\substack{k=1 \\ k \neq j}}^{n_p} d_{i,k} \left([\pi_k] - \frac{\beta_{i,k}}{d_{i,k}} \right)^2 - \gamma_j + \sum_{k=1}^{n_p} \frac{\beta_{i,k}^2}{d_{i,k}} \right) \right)^{\frac{1}{2}} + \frac{\beta_{i,j}}{d_{i,j}} \right\}$$
$$\tag{36}$$
$$\in \left[\pi'_j \right],$$

which completes the proof. $\diamond$

Applying the contractors $\mathcal{C}_{z_i - z_0, [0, \infty[}$, $i = 1, \ldots, m$, to $[\mathbf{p}]$, as in Section 3.1.1, one obtains a list of m possibly contracted boxes $\mathcal{L} = \left\{ \mathcal{C}_{z_1 - z_0, [0, \infty[}([\mathbf{p}]), \ldots, \mathcal{C}_{z_m - z_0, [0, \infty[}([\mathbf{p}]) \right\}$.

3.1.3 Building a q-relaxed intersection

During the second step, the contractor builds a box $[\mathbf{p}']$ enclosing the q-relaxed intersection $\mathcal{P}$ [5,6] 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 $\mathcal{L}$

$$\mathcal{P} = \bigcap_{j \in \{1, \ldots, m-1\}}^{q} [\mathbf{p}'_j] = \bigcup_{\substack{J \subset [1, \ldots, m-1] \\ \text{card}(J) \geqslant q}} \bigcap_{j \in J} [\mathbf{p}'_j], \quad (37)$$

with

$$\mathcal{P} \subset [\mathbf{p}'] \subset [\mathbf{p}]. \tag{38}$$

Proposition 2 *For any box $[\mathbf{p}']$ satisfying (38),*

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

with $\boldsymbol{\Psi}_q$ as defined in (14).

---

1    $[p] = \emptyset$;

2   Reindex the boxes $[p_i]$ in such a way that
$$\underline{p}_1 \leqslant \underline{p}_2 \leqslant \cdots \leqslant \underline{p}_n;$$

3   For $i = q$ to n

4   if $\sum_{j=1}^{n} \left(\underline{p}_i \in [p_j] \right) \geqslant q$

5        $\underline{p} = \underline{p}_i$; break;

6   Reindex the boxes $[p_i]$ in such a way that
$$\overline{p}_1 \geqslant \overline{p}_2 \geqslant \cdots \geqslant \overline{p}_n$$

7   For $i = q$ to n

8   if $\sum_{j=1}^{n} \left(\overline{p}_i \in [p_j] \right) \geqslant q$

9        $\overline{p} = \overline{p}_i$;

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

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 (14), $\sum_{i=1}^{m} \tau_i(\mathbf{p}_0) \geqslant q$. 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. $\diamond$

3.1.4 Evaluating the q-relaxed intersection

Algorithm 1 formalizes a computation carried out on an example in [6]. It aims at building an outer approximating interval of the q-relaxed intersection of m scalar intervals.

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. The extension to boxes is obtained by applying Algorithm 1 componentwise.

4 Examples

A model the output of which is nonlinear in its parameters is considered first with LSCR. Then, a FIR model with a large number of parameters is considered with SPS.

All computations were carried out with Intlab [10], the interval-analysis toolbox for Matlab, on an Intel Core i7 at 3.7 GHz with 8 GB RAM. The computations required by SIVIA and the q-relaxed intersection, which form the major part of the computational burden, could
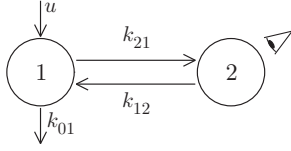
5

Fig. 1. Two-compartment model

be speeded up considerably with a C++ implementation.

4.1 Model nonlinear in its parameters

Consider the two-compartment model described by Figure 1. Only the content of the second compartment is observed. The model parameters to be estimated are $\mathbf{p} = (k_{01}, k_{12}, k_{21})^{\mathrm{T}}$. The data are generated by this model for some true value $\mathbf{p}^* = (1, 0.25, 0.5)^{\mathrm{T}}$ of its parameter vector. They satisfy

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

where

$$\alpha\left(\mathbf{p}\right) = k_{21}/\sqrt{\left(k_{01} - k_{12} + k_{21}\right)^2 + 4k_{12}k_{21}}, \quad (39)$$

$$\lambda_{1,2}\left(\mathbf{p}\right) = -\frac{1}{2}\left(\left(k_{01} + k_{12} + k_{21}\right)\right.$$
$$\left.\pm\left(\left(k_{01} - k_{12} + k_{21}\right)^2 + 4k_{12}k_{21}\right)^{-1/2}\right) \quad (40)$$

and the w_t's are realizations of i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$ variables, for $t = 0, T, \ldots, (n-1)T$. 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. The prediction errors are $\varepsilon_t\left(\mathbf{p}\right) = y_t - y_t^{\mathrm{m}}\left(\mathbf{p}\right)$, with

$$y_t^{\mathrm{m}}\left(\mathbf{p}\right) = \alpha\left(\mathbf{p}\right)\left(\exp\left(\lambda_1\left(\mathbf{p}\right)t\right) - \exp\left(\lambda_2\left(\mathbf{p}\right)t\right)\right),$$

for $t = 0, T, \ldots, (n-1)T$.

Here, the set $\mathbf{\Theta}_{r,q}^\varepsilon$ has been characterized using LSCR with $r = 1$ and $q = 3$, which corresponds to a 90 % confidence region, see Figure 2. The initial search set in parameter space is $\mathbb{P} = [0,5] \times [0,5] \times [0.5, 0.5]$. The top left part of Figure 2 represents the result obtained in 284s by gridding as in [1] with a grid step-size $\varepsilon = 0.0025$. The top right part of Figure 2 has been obtained by SIVIA with $\varepsilon = 0.0025$ in 175s. The top right part of Figure 2 *proves* that the confidence region consists of two disconnected subsets, a consequence of the lack of global identifiability of the model (the values of k_{01} and k_{12} may be exchanged without changing the model output). Figure 2 (bottom part) zooms on one of the two confidence subsets, which turns out to contain the
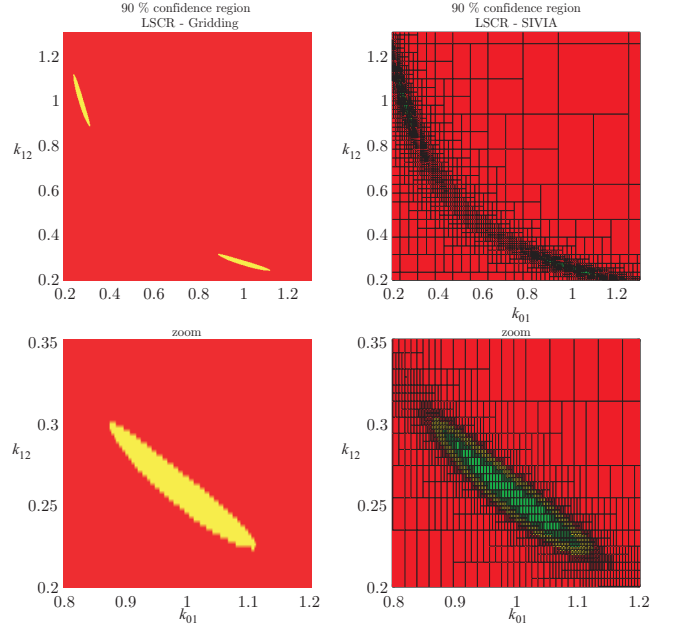

Fig. 2. Gridding (left) and paving (right) of search space obtained using LSCR in the two-compartment model case for the characterization of the confidence region $\mathbf{\Theta}_{r,q}^\varepsilon$; the two bottom subfigures are obtained by zooming in on one of the connected components of $\mathbf{\Theta}_{r,q}^\varepsilon$

actual value of the unknown parameters, although this is not guaranteed, of course.

Table 1 shows the evolution of computing time for the gridding approach and SIVIA. The increase is quadratic with $1/\varepsilon$ for gridding. It is slower with SIVIA, since only undetermined boxes are further bisected when ε decreases. An additional advantage of SIVIA is that the results it provides are guaranteed.

ε	0.1	0.025	0.01	0.0025	0.001
Gridding (s)	0.26	2.9	18	284	1750
SIVIA (s)	15	57	93	175	400

Table 1
Computing times for various values of ε in the example of Section 4.1

4.2 FIR model

Consider now the system

$$y_t = y_t^{\mathrm{m}}\left(\mathbf{p}^*\right) + w_t, \quad (41)$$

with the FIR model

$$y_t^{\mathrm{m}}\left(\mathbf{p}\right) = \sum_{i=0}^{n_{\mathrm{a}}-1} a_i u_{t-i}, \quad (42)$$

6

where $\mathbf{p} = (a_0, \ldots, a_{n_\mathrm{a}-1})^\mathrm{T}$ and $u_t = 0$ for $t \leqslant 0$. For $t = 1, \ldots, n$, the w_t's are i.i.d. noise samples. In linear regression form, (41) becomes

$$y_t = \boldsymbol{\varphi}_t^\mathrm{T} \mathbf{p}^* + w_t \qquad (43)$$

with $\boldsymbol{\varphi}_t^\mathrm{T} = (u_t, \ldots, u_{t-n_\mathrm{a}+1})$ and $\mathbf{p}^* = (a_0^*, \ldots, a_{n_\mathrm{a}-1}^*)^\mathrm{T}$.
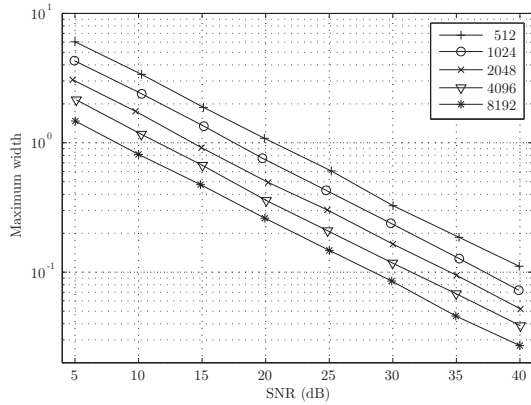
Fig. 3. Width of the largest component of the outer box resulting from a single application of the contractor of Section 3.1.2 as a function of the SNR and number of data points.

To evaluate the performance of the proposed technique for a large number of parameters, FIR models (42) with $n_\mathrm{a} = 20$ random parameters in $[-2, 2]^{n_\mathrm{a}}$ are generated. Then, $n = 512, 1024, 2048, 4096$, and 8192 noise-free data points are generated applying a random i.i.d. sequence u_t of ± 1, which is the D-optimal input under the constraint that the input has to remain in $[-1, 1]$ [11]. White Laplacian noise is then added to these data. The standard deviation of the noise is set up to get SNRs from 5 dB to 40 dB.

Our aim is to characterize a $95\,\%$ confidence region with SPS. A possible choice is $m = 255$ and $q = 13$. The initial search box in parameter space is taken as $\mathbb{P} = [-10^4, 10^4]^{20}$. Getting accurate inner and outer approximations using union of non-overlapping boxes is hopeless with $n_\mathrm{a} = 20$ parameters, because of the curse of dimensionality. Our aim is instead to provide a guaranteed outer-approximation of the confidence region[1]. For that purpose, the contractor of Section 3.1.2 is applied once to $\mathbb{P}$ (iterations are useless). A box estimate is obtained in 5s in average, whatever the value of n. This is because the computational complexity of the contractor of Section 3.1.2 is mainly determined by m and n_a. Figures 3 and 4 represent the width of the largest component of the resulting outer box as a function of the SNR and of the number of data points.

---

[1] The computed least-square estimate belongs to the confidence region, as showed in [2]. It thus forms a (point) inner approximation.
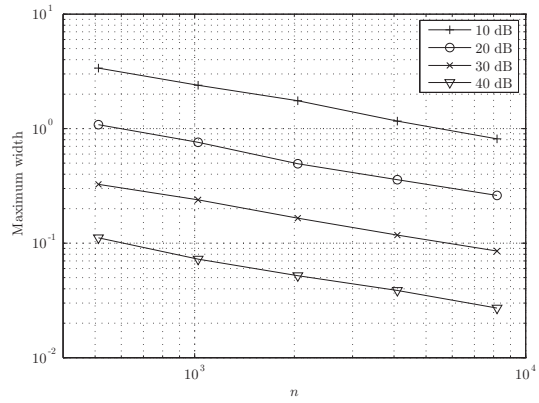

Fig. 4. Width of the largest component of the outer box resulting from a single application of the contractor of Section 3.1.2 as a function of the number of data points and of the SNR.

The width of the box decreases linearly (in the log domain) when the SNR or the log of the number of data points increases.

5   Conclusions and perspectives

Interval analysis provides tools to evaluate guaranteed inner and outer-approximations of non-asymptotic confidence regions defined by LSCR and SPS. This has been demonstrated on a model the output of which is nonlinear in its parameters.

Accurate inclusion functions are particularly difficult to obtain for the functions involved in LSCR or SPS, due to the many occurrences of the parameters involved in the evaluation of (3) for LSCR and of (9) and (10) for SPS. 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 more efficient contractors than the one considered in Section 3.1.1. This is what was done in Section 3.1.2 for the linear case for SPS. It then becomes possible to get guaranteed outer-approximations of confidence regions defined by SPS for FIR models with a large number of parameters.

References

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

[2] B. C. Csáji, M. C. Campi, and E. Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proc. SYSID*, pages 227–232, Brussels, Belgium, 2012.

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

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

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

[6] L. Jaulin. Set-membership localization with probabilistic errors. *Rob. and Auton. Syst.*, 59(6):489–495, 2011.

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

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

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

[10] 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.

[11] E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Springer, London, 1997.