Bayesian estimation using interval analysis

Luc Jaulin

Laboratoire E^3I^2

(Extraction et Exploitation de l'Information en Environnements Incertains)

ENSIETA, 2 rue François Verny

29806 Brest Cédex 09

E-mail : luc.jaulin@ensieta.fr

1. Introduction

Parameter set estimation deals with characterizing a (preferably small) set which encloses the actual parameter vector $\mathbf{p}^*$ of a parametric model from measurement data. In the context of bounded-error estimation, the measurement error is assumed to be bounded and the feasible set for the parameter vector $\mathbf{p}$ can be written [6] as the solution set of a constraint satisfaction problem (CSP) for which interval constraint propagation methods have been shown to be particularly efficient [1]. In a probabilistic context, the error is not anymore described by membership intervals, but by *probability density functions* (pdf). The Bayes rule then makes it possible to obtain the posterior[1] pdf for $\mathbf{p}$ (see, e.g., [2]). The set estimation in such a probabilistic context is often called *Bayesian estimation*. The set to be estimated becomes the *optimal confidence region* and corresponds to the smallest set $\mathbb{S}_\alpha$, in the parameter space, which contains $\mathbf{p}^*$ with a given probability α. This set cannot be described by a CSP and existing interval methods cannot be used without a serious adaptation. To my knowledge, the problem of estimating optimal confidence regions has never been solved before in its general form (see [4] where it is assumed that an analytical expression of the distribution function of the pdf is available, which is not realistic).

The goal of this paper is to show that interval methods can be used to deal with set estimation even in a probabilistic context (see also [8], for interval techniques used in a probabilistic context). This could help to make interval analysis more attractive to the Bayesian estimation community, most of who almost ignore interval computation, except maybe for its ability to deal with global optimization in a guaranteed way [3].

Section 2 describes in details the problem of characterizing optimal confidence regions of a pdf. Some notions on intervals, subpavings and staircase functions, needed to understand the methodology, are given in Section 3. These notions will make it possible to solve our problem in Section 4. An application to Bayesian estimation is then considered in Section 5.

[1] The prior (resp. posterior) pdf is the pdf for $\mathbf{p}$ before (resp. after) taking into account the measured data.

2. Definition of the problem

In Bayesian estimation, we are generally able to write the posterior probability density function (pdf) for the parameter vector $\mathbf{p} \in \mathbb{R}^n$ to be estimated into the form

$$\pi(\mathbf{p}) = \frac{1}{a} f(\mathbf{p}), \tag{2.1}$$

where a is a real number and f is an integrable positive function for which an analytical expression is available (see Section 5 for an illustrative example). When $a \neq 1$ the function f is said to be an *unnormalized pdf.* In general a is not known but can be computed approximately by taking into account that π is a pdf:

$$\int_{\mathbb{R}^n} \pi(\mathbf{p}) d\mathbf{p} = 1, \tag{2.2}$$

i.e.,

$$a = \int_{\mathbb{R}^n} f(\mathbf{p}) d\mathbf{p}. \tag{2.3}$$

The *optimal confidence region* $\mathbb{S}_\alpha$ of level α associated with $\pi(\mathbf{p})$ is the minimum volume set which contains $\mathbf{p}$ with a probability equal to α. Except for atypical cases that might appear when π has some flat levels, the set $\mathbb{S}_\alpha$ can be expressed into a set inversion form as

$$\mathbb{S}_\alpha = f^{-1}([s_\alpha, +\infty]), \tag{2.4}$$

where $s_\alpha \in \mathbb{R}$ is the *threshold* associated with the probability α. The threshold s_α is unknown, but should satisfy

$$\int_{\mathbb{S}_\alpha} \pi(\mathbf{p}) d\mathbf{p} = \alpha \tag{2.5}$$

or equivalently

$$\frac{1}{a} \int_{\mathbb{S}_\alpha} f(\mathbf{p}) d\mathbf{p} = \alpha. \tag{2.6}$$

The problem to be considered in this paper is the characterization of $\mathbb{S}_\alpha$ for any given $\alpha \in [0, 1]$. It can be formulated as follows:

Problem: Consider a function $f(\mathbf{p})$ positive for all $\mathbf{p}$, such as $\int_{\mathbb{R}^n} f(\mathbf{p}) d\mathbf{p}$ is finite and a real number $\alpha \in [0, 1]$. Characterize the set $\mathbb{S}_\alpha$ defined by

$$\begin{array}{ll} \text{(i)} & \mathbb{S}_\alpha = f^{-1}([s_\alpha, +\infty[), \\ \text{(ii)} & \frac{\int_{\mathbb{S}_\alpha} f(\mathbf{p}) d\mathbf{p}}{\int_{\mathbb{R}^n} f(\mathbf{p}) d\mathbf{p}} = \alpha. \end{array} \tag{2.7}$$

The relation (ii) is a consequence of equations (2.3) and (2.6).

The following example illustrates these notions and shows the difficulty to characterize $\mathbb{S}_\alpha$ even for academic functions f.

Example 2.1. *Consider a random variable* p*, described by a* standard normal pdf $\pi(p) = \frac{1}{a}f(p)$ *where* $f(p) = \exp\left(-\frac{p^2}{2}\right)$*.Let us compute its optimal confidence region* $\mathbb{S}_{0.95}$ *of level* 0.95*. From Equation (2.7, i),* $\mathbb{S}_{0.95}$ *is an interval of the form* $\mathbb{S}_{0.95} = [-b, b]$*. Equation (2.7, ii) translates into*

$$\int_{-b}^{b} \exp\left(-\frac{p^2}{2}\right) dp = \alpha. \int_{-\infty}^{\infty} \exp\left(-\frac{p^2}{2}\right) dp \tag{2.8}$$

Since, $\int_{-\infty}^{\infty} \exp\left(-\frac{p^2}{2}\right) dp = \sqrt{2\pi}$and $\int_{-b}^{b} \exp\left(-\frac{p^2}{2}\right) dp = \sqrt{2\pi}\,\mathrm{erf}\left(\frac{1}{2}b\sqrt{2}\right)$ *, where* erf *is the well-known* error function, *the equation (2.8) becomes* $\sqrt{2\pi}\,\mathrm{erf}\left(\frac{1}{2}b\sqrt{2}\right) = \alpha.\sqrt{2\pi}$*.By taking into account the fact that* erf *is an increasing function, for* $\alpha = 0.95$*, we get* $b = 1.96$ *using any local numerical method. The threshold is* $s_\alpha = f(b) = \exp\left(-\frac{1.96^2}{2}\right) = 0.1465$*. We can also write* $\mathbb{S}_{0.95} = f^{-1}([0.1465, \infty[) = [-1.96, 1.96]$*.. Figure 2.1 gives an illustration of this result. The grey area contains 95% of all the area below* $f(p)$*. Let us stress that one cannot find any subset* $\mathbb{S}'$ *smaller (*i.e., *with a smaller length) than* $\mathbb{S}_\alpha$ *such that* $\int_{\mathbb{S}'} f(p)dp \geq 0.95$*.*

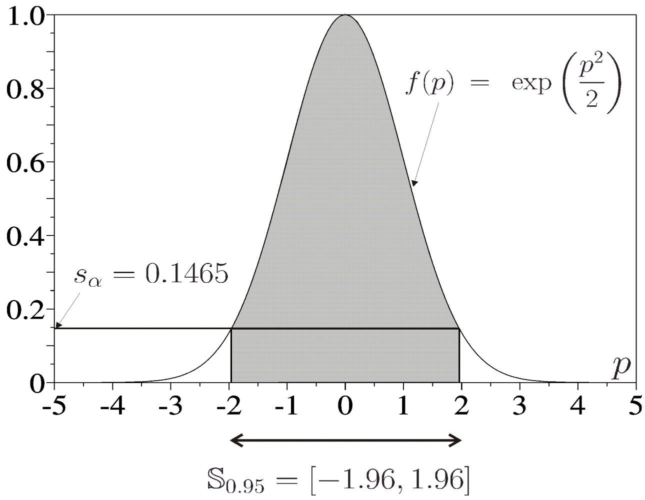

Figure 2.1: Optimal confidence region associated with the standard normal probability density function; the grey area contains 95% of all the area below the bell-shaped function.

3. Intervals

3.1. Intervals of $\mathbb{R}$ and boxes

A *lattice* $(\mathcal{E}, \leq)$ is a partially ordered set, closed under least upper and greatest lower bounds. The least upper bound (*join*) of x and y is written $x \vee y$. The greatest lower bound (*meet*) is written $x \wedge y$. A lattice $\mathcal{E}$ is *complete* if for all (finite or not) subsets $\mathcal{A}$ of $\mathcal{E}$, the least upper bound of $\mathcal{A}$ and the greatest lower bound of $\mathcal{A}$ belong to $\mathcal{E}$. An *interval* $[x]$ of a complete lattice $\mathcal{E}$ is a subset of $\mathcal{E}$ which satisfies

$$[x] = \{x \in \mathcal{E} \mid \wedge [x] \leq x \leq \vee [x]\}. \tag{3.1}$$

Note that both $\emptyset$ and $\mathcal{E}$ are intervals of $\mathcal{E}$.

Example 3.1. *The set* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ *is a complete lattice. The intervals of* $\overline{\mathbb{R}}$ *are the closed connected subsets of* $\bar{\mathbb{R}}$*. The set of all intervals of* $\bar{\mathbb{R}}$ *will be denoted by* $\mathbb{I}\bar{\mathbb{R}}$*.*

Example 3.2. *The set* $\bar{\mathbb{R}}^n$ *is a complete lattice with respect to the partial order relation given by* $\mathbf{x} \leq \mathbf{y} \Leftrightarrow \forall i \in \{1, \ldots, n\}, x_i \leq y_i$*. We have* $\mathbf{x} \wedge \mathbf{y} = (\min(x_1, y_1), \ldots, \min(x_n, y_n))$ *and* $\mathbf{x} \vee \mathbf{y} = (\max(x_1, y_1), \ldots, \max(x_n, y_n))$*. The intervals of* $\bar{\mathbb{R}}^n$ *are called boxes. The set of all boxes of* $\bar{\mathbb{R}}^n$ *will be denoted by* $\mathbb{I}\bar{\mathbb{R}}^n$*.*

3.2. Interval subpavings

A *paving* $\mathcal{Q}$ of $\mathbb{R}^n$ is a finite set of non overlapping boxes covering $\mathbb{R}^n$. A *subpaving* of $\mathcal{Q}$ is a subset of $\mathcal{Q}$. The set of all subpavings of $\mathcal{Q}$ will be denoted by $\mathcal{P}(\mathcal{Q})$. One can show that $(\mathcal{P}(\mathcal{Q}), \subset)$ is a complete lattice. The least upper bound (*join*) is the union: $\mathcal{K}_1 \vee \mathcal{K}_2 = \mathcal{K}_1 \cup \mathcal{K}_2$ and the greatest lower bound (*meet*) is the intersection $\mathcal{K}_1 \wedge \mathcal{K}_2 = \mathcal{K}_1 \cap \mathcal{K}_2$. As a consequence intervals of $(\mathcal{P}(\mathcal{Q}), \subset)$ can be defined. An interval subpaving of $\mathcal{Q}$ can be represented by a pair of two subpavings $[\mathcal{K}^-, \mathcal{K}^+]$ of $\mathcal{Q}$ such that $\mathcal{K}^- \subset \mathcal{K}^+$, as illustrated by Figure 3.1.

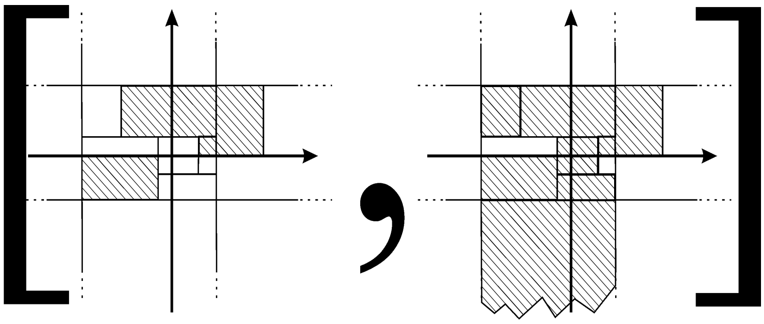

Figure 3.1: An interval subpaving is a pair of two subpavings $[\mathcal{K}^-, \mathcal{K}^+]$ of $\mathcal{Q}$ such that $\mathcal{K}^- \subset \mathcal{K}^+$

Example 3.3. *Consider the paving of* $\mathbb{R}$ *defined by* $\mathcal{Q} = \{]-\infty, 0], [0, 1], [1, 2], [2, 3], [3, \infty[\}$*. Three possible interval subpavings of* $\mathcal{Q}$ *are*

$$
\begin{aligned}
[\mathcal{K}_1^-; \mathcal{K}_1^+] &= [\{]-\infty, 0], [1, 2]\}; \{]-\infty, 0], [1, 2], [2, 3]\}], && (3.2)\\
[\mathcal{K}_2^-; \mathcal{K}_2^+] &= [\{[1, 2]\}; \{[1, 2]\}], && (3.3)\\
[\mathcal{K}_3^-; \mathcal{K}_3^+] &= [\{\}; \{\}]. && (3.4)
\end{aligned}
$$

The *support* $\{\mathcal{K}\} \subset \mathbb{R}^n$ of a subpaving $\mathcal{K} \in \mathcal{P}(\mathcal{Q})$ is the union of all boxes of $\mathcal{K}$ (see Figure 3.2). A subpaving $\mathcal{K}$ is said to be *equivalent* to a subset $\mathbb{S}$ of $\mathbb{R}^n$ (we shall write $\mathcal{K} \equiv \mathbb{S}$) if $\{\mathcal{K}\} = \mathbb{S}$. A set $\mathbb{S}$ is said to belong to the interval subpaving $[\mathcal{K}^-, \mathcal{K}^+]$ if all boxes of $\mathcal{K}^-$ are included in $\mathbb{S}$ and if all elements of $\mathbb{S}$ belong to at least one box of $\mathcal{K}^+$, *i.e.*,

$$
\mathbb{S} \in [\mathcal{K}^-, \mathcal{K}^+] \Leftrightarrow \{\mathcal{K}^-\} \subset \mathbb{S} \subset \{\mathcal{K}^+\}. \quad (3.5)
$$

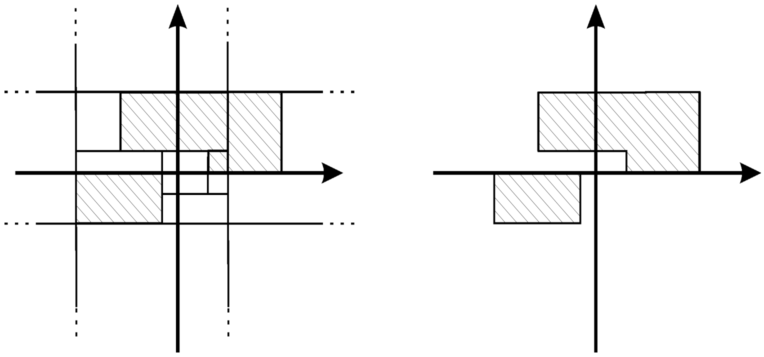

Figure 3.2: A subpaving $\mathcal{K}$ (left) and its support $\{\mathcal{K}\}$ (right)

3.3. Interval staircase functions

A *staircase function* $\hat{f}$ associated with a paving $\mathcal{Q}$ is a function from $\mathcal{Q}$ to $\bar{\mathbb{R}}$. The set of all staircase functions $(\hat{\mathcal{F}}, \leq)$ is a complete lattice. Interval staircase functions can thus be defined. An interval staircase function $[\hat{f}] = [\hat{f}^-, \hat{f}^+]$ can be represented a pair of two staircase functions such that $\forall [\mathbf{p}] \in \mathcal{Q}\ ,\ \hat{f}^-([\mathbf{p}]) \leq \hat{f}^+([\mathbf{p}])$.Since $\mathcal{Q}$ is finite, $[\hat{f}]$ can easily be represented in the computer. Note that interval staircase functions have already been used to approximate pdf in [8]. A function f from $\mathbb{R}^n \to \mathbb{R}$ is said to belong to the interval staircase function $[\hat{f}]$ if

$$\forall [\mathbf{p}] \in \mathcal{Q}, \forall \mathbf{p} \in [\mathbf{p}], f(\mathbf{p}) \in [\hat{f}^-([\mathbf{p}]), \hat{f}^+([\mathbf{p}])]. \tag{3.6}$$

An interval staircase function for a function $f : \mathbb{R}^n \to \mathbb{R}$ can easily be obtained by using interval techniques as described in [7].

The *reciprocal image* of the interval $[s^-, s^+] \in \mathbb{IR}$ by the interval staircase function $[\hat{f}] = [\hat{f}^-, \hat{f}^+]$ is the interval subpaving of $\mathcal{Q}$ defined by

$$[\hat{f}]^{-1}([s^-, s^+]) \triangleq \left[\left\{ [\mathbf{p}] \in \mathcal{Q} \mid [\hat{f}]([\mathbf{p}]) \subset [s^-, s^+] \right\}, \left\{ [\mathbf{p}] \in \mathcal{Q} \mid [\hat{f}]([\mathbf{p}]) \cap [s^-, s^+] \neq \emptyset \right\} \right]$$

Theorem 3.4. *If f belongs to $[\hat{f}]$, then for all interval $[s^-, s^+] \in \mathbb{I}\bar{\mathbb{R}}$, then $f^{-1}([s^-, s^+]) \in [\hat{f}]^{-1}([s^-, s^+])$.*

Proof: Trivial.

Example 3.5. *Figure 3.3 illustrates this theorem for $[s^-, s^+] = [16, \infty]$ and $\mathcal{Q} = \{[i, i+1],\ i \in \mathbb{N}\}$. We have*

$$\{[\mathbf{p}] \in \mathcal{Q} \mid [\hat{f}]([\mathbf{p}]) \subset [s^-, s^+]\ \} = \{[-1, 0], [0, 1]\} \equiv [-1, 1],$$
$$\{[\mathbf{p}] \in \mathcal{Q} \mid [\hat{f}]([\mathbf{p}]) \cap [s^-, s^+] \neq \emptyset\} = \{[-3, -2], [-2, -1], [-1, 0], [0, 1], [1, 2], [2, 3]\} \equiv [-3, 3].$$

and thus

$$[-1, 1] \subset f^{-1}([16, \infty]) \subset [-3, 3].$$

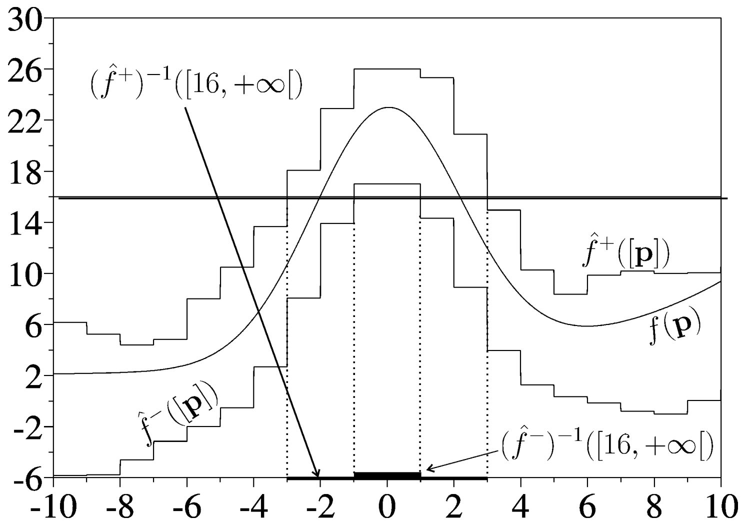

Figure 3.3: An interval staircase function enclosing a function $f(\mathbf{p})$

If $[\mathcal{K}^-,\mathcal{K}^+]$ is an interval subpaving of $\mathcal{Q}$ and if $[\hat{f}]$ is an interval staircase function, we define the *integral* of $[\hat{f}]$ over $[\mathcal{K}^-,\mathcal{K}^+]$ by

$$\int_{[\mathcal{K}^-,\mathcal{K}^+]} [\hat{f}](\mathbf{p})d\mathbf{p} = \left[\sum_{[\mathbf{p}]\in\mathcal{K}^-} \hat{f}^-([\mathbf{p}]).\text{volume}([\mathbf{p}]), \sum_{[\mathbf{p}]\in\mathcal{K}^+} \hat{f}^+([\mathbf{p}]).\text{volume}([\mathbf{p}])\right]. \tag{3.7}$$

Theorem 3.6. *If f is a positive function which belongs to the interval staircase function $[\hat{f}]$, if $\mathbb{S} \in [\mathcal{K}^-,\mathcal{K}^+]$ is a measurable set then*

$$\int_{\mathbb{S}} f(\mathbf{p})d\mathbf{p} \in \int_{[\mathcal{K}^-,\mathcal{K}^+]} [\hat{f}](\mathbf{p})d\mathbf{p}. \tag{3.8}$$

Proof:

$$\begin{aligned}
\int_{\mathbb{S}} f(\mathbf{p})d\mathbf{p} &\in \left[\int_{\{\mathcal{K}^-\}} f(\mathbf{p})d\mathbf{p}, \int_{\{\mathcal{K}^+\}} f(\mathbf{p})d\mathbf{p}\right] && (3.9)\\
&= \left[\sum_{[\mathbf{p}]\in\mathcal{K}^-} \int_{[\mathbf{p}]} f(\mathbf{p})d\mathbf{p}, \sum_{[\mathbf{p}]\in\mathcal{K}^+} \int_{[\mathbf{p}]} f(\mathbf{p})d\mathbf{p}\right] && (3.10)\\
&\subset \left[\sum_{[\mathbf{p}]\in\mathcal{K}^-} \int_{[\mathbf{p}]} \hat{f}^-(\mathbf{p})d\mathbf{p}, \sum_{[\mathbf{p}]\in\mathcal{K}^+} \int_{[\mathbf{p}]} f^+(\mathbf{p})d\mathbf{p}\right] && (3.11)\\
&= \left[\sum_{[\mathbf{p}]\in\mathcal{K}^-} \hat{f}^-([\mathbf{p}]).\text{volume}([\mathbf{p}]), \sum_{[\mathbf{p}]\in\mathcal{K}^+} \hat{f}^+([\mathbf{p}]).\text{volume}([\mathbf{p}])\right] && (3.12)\\
&= \int_{[\mathcal{K}^-,\mathcal{K}^+]} [\hat{f}](\mathbf{p})d\mathbf{p}. && (3.13)
\end{aligned}$$

4. Algorithm

In this section, we shall provide a numerical interval algorithm to characterize optimal confidence region of level α of an unnormalized pdf $f(\mathbf{p})$. To solve this problem we need to solve the following equation (see (2.7)):

$$\alpha = h(s_\alpha) \triangleq \frac{\int_{f^{-1}([s_\alpha,\infty[)} f(\mathbf{p})d\mathbf{p}}{\int_{\mathbb{R}^n} f(\mathbf{p})d\mathbf{p}} \tag{4.1}$$

where the single unknown is s_α. Note that, since

$$\begin{aligned} s_1 \quad &\leq \quad s_2 \Rightarrow [s_2,\infty[\subset [s_1,\infty[\Rightarrow f^{-1}([s_2,\infty[) \subset f^{-1}([s_1,\infty[) & (4.2)\\ &\Rightarrow \int_{f^{-1}([s_2,\infty[)} f(\mathbf{p})d\mathbf{p} \leq \int_{f^{-1}([s_1,\infty[)} f(\mathbf{p})d\mathbf{p}, & (4.3) \end{aligned}$$

the function $h(s)$ is decreasing. Moreover,

$$h(s) \in [h](s) \triangleq \frac{\int_{[\hat{f}]^{-1}([s,\infty[)} [f](\mathbf{p})d\mathbf{p}}{\int_{\mathbb{R}^n} [\hat{f}](\mathbf{p})d\mathbf{p}}. \tag{4.4}$$

where $[h](s)$ is a function from $\mathbb{R} \rightarrow \mathbb{IR}$.

Remark 1. *For a given s, to compute $[h](s)$, one should first compute the interval subpaving $[\mathcal{K}^-,\mathcal{K}^+] = [\hat{f}]^{-1}([s,\infty[)$ (see Equation (**??**)). Then, we have to compute the two intervals $\int_{[\hat{f}]^{-1}([s,\infty[)} [\hat{f}](\mathbf{p})d\mathbf{p}$ and $\int_{\mathbb{R}^n} [\hat{f}](\mathbf{p})d\mathbf{p}$ (see Equation (3.7)). The two resulting intervals are then divided using interval arithmetic rules.*

Since $h(s)$ is decreasing and since $h(s) \in [h](s)$, we have the following implications

$$\begin{array}{lll} (a) & \alpha < lb([h](s^-)) & \Rightarrow s^- < s_\alpha \\ (b) & \alpha > ub([h](s^+)) & \Rightarrow s^+ > s_\alpha \end{array} \tag{4.5}$$

where lb and ub correspond to the lower and upper bounds, respectively. These two implications are illustrated by Figure 4.1.

The principle of the method to compute an enclosure $[s^-, s^+]$ of s_α is to decrease the value s^- initialized at $+\infty$ until $\alpha < lb([h](s^-))$ and to increase the value s^+ initialized at 0 until $\alpha > ub([h](s^+))$.

The procedure for computing $\mathbb{S}_\alpha$ is as follows:

1. Take a paving $\mathcal{Q}$ of $\mathbb{R}^n$;
2. Compute an interval staircase function $[\hat{f}]$ enclosing f;
3. Define the two lists $\hat{f}^-(\mathcal{Q}) := \left\{\hat{f}^-([\mathbf{p}]), [\mathbf{p}] \in \mathcal{Q}\right\}$ and $\hat{f}^+(\mathcal{Q}) := \left\{\hat{f}^+([\mathbf{p}]), [\mathbf{p}] \in \mathcal{Q}\right\}$;

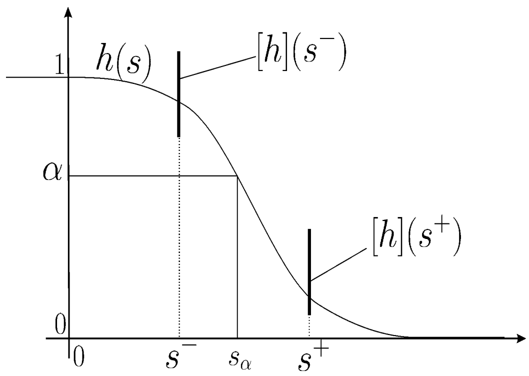

Figure 4.1: Illustration of the two tests used to find an upper and a lower bound for s_α

4. Denote by s^- the largest element of $\hat{f}^-(\mathcal{Q})$;
5. If $\alpha \geq lb([h](s^-))$, remove s^- from the list $\hat{f}^-(\mathcal{Q})$ and go to 4;
6. Denote by s^+ the smallest element of $\hat{f}^+(\mathcal{Q})$;
7. If $\alpha \leq ub([h](s^+))$, remove s^+ from the list $\hat{f}^+(\mathcal{Q})$ and go to 6;
8. $[\mathcal{K}_\alpha^-, \mathcal{K}_\alpha^+] := ([\hat{f}] - [s^-, s^+])^{-1}([0, \infty[)$.

Comments: At Step 1 a finite paving $\mathcal{Q}$ is chosen in order to have small boxes where f varies significantly and big boxes where the density is almost zero. The interval staircase function $[\hat{f}]$, computed at Step 2, is obtained using interval arithmetic and is then stored in a list whose elements are pairs $\left([\mathbf{p}]_i, [\hat{f}]_i\right)$ where $[\mathbf{p}]_i$ is the ith box and $[\hat{f}]_i$ corresponds to $[\hat{f}]([\mathbf{p}])$. At Step 3, we define the lists $\hat{f}^-(\mathcal{Q})$ and $\hat{f}^+(\mathcal{Q})$ which correspond to all values of s where $lb([h](s))$ and $ub([h](s))$ can change, respectively. Thus, the values for s^- and s^+ will all be taken in $\hat{f}^-(\mathcal{Q})$ and $\hat{f}^+(\mathcal{Q})$, respectively. Steps 4 and 5 compute a lower bound s^- for s_α. Steps 6 and 7 compute an upper bound s^+ for s_α. Recall that $[h](s^-)$ and $[h](s^+)$ are computed using the procedure described on Remark 1. When the algorithm reaches Step 8, we have $\alpha < lb([h](s^-))$ and $\alpha > ub([h](s^+))$ and thus $s_\alpha \in [s^-, s^+]$. At Step 8, when we compute $[\hat{f}] - [s^-, s^+]$, we use abusively of an arithmetic over interval staircase functions (such an arithmetic is not be defined in this paper). In fact, $[\hat{f}] - [s^-, s^+]$ denotes the interval staircase function associated with the function $f - s_\alpha$. Indeed, $\mathbb{S}_\alpha = f^{-1}([s_\alpha, \infty[) = (f - s_\alpha)^{-1}([0, \infty[)$ and thus from Theorem 3.4, $\mathbb{S}_\alpha \in [\mathcal{K}_\alpha^-, \mathcal{K}_\alpha^+]$.

Theorem 4.1. *After completion of this algorithm, we have*

$$\mathbb{S}_\alpha \in \left[\mathcal{K}_\alpha^-, \mathcal{K}_\alpha^+\right] \text{ and } s_\alpha \in [s^-, s^+]. \tag{4.6}$$

A solver implementing this algorithm can be freely downloaded with its C++ Builder 5 code [5].

5. Application to Bayesian estimation

Consider now a parametric model for a system for which m output data have been collected (see [9] for notions related to parameter estimation). All these data have been stored within a vector $\mathbf{y}$. The output vector $\mathbf{y}$ is assumed to be related to the parameter vector by

$$\mathbf{y} = \phi(\mathbf{p}) + \mathbf{n} \tag{5.1}$$

where $\mathbf{n}$ is the noise vector, and ϕ is the model function. In a probabilistic approach, we generally assume that a prior pdf $\pi_{\text{prior}}(\mathbf{p})$ for $\mathbf{p}$ and $\pi_n(\mathbf{n})$ for $\mathbf{n}$ are known. The Bayes rule (see *e.g.* [2], [9]) states that the posterior pdf for $\mathbf{p}$ is given by

$$\pi_{\text{post}}(\mathbf{p}) = \frac{\pi_n(\mathbf{y} - \phi(\mathbf{p})).\pi_{\text{prior}}(\mathbf{p})}{\int_{\mathbf{p}\in\mathbb{R}^n} \pi_n(\mathbf{y} - \phi(\mathbf{p})).\pi_{\text{prior}}(\mathbf{p})d\mathbf{p}}. \tag{5.2}$$

For most applications, we have an analytical expression for $\pi_n(\mathbf{n})$ and $\pi_{\text{prior}}(\mathbf{p})$, but not [2] for the denominator a of (5.2). Therefore, characterizing the optimal confidence region of $\pi_{\text{post}}(\mathbf{p})$ amounts to solving (2.7) where the unnormalized pdf is

$$f(\mathbf{p}) = \pi_n(\mathbf{y} - \phi(\mathbf{p})).\pi_{\text{prior}}(\mathbf{p}). \tag{5.3}$$

As an illustration, consider the model described by

$$y(t) = p_1 \sin(p_2 t) + n(t) \tag{5.4}$$

where $n(t)$ is a white random signal (for all t_1, t_2, the random variables $n(t_1)$ and $n(t_2)$ are independent). For each t, its probability density function is normal, *i.e.*, it satisfies

$$\pi_n(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n^2}{2\sigma^2}\right) \tag{5.5}$$

where the standard deviation σ will be taken equal to $\frac{1}{2}$. Assume that at sampling times $\mathbf{t} = (1, 2, 3)$, the three following measurements have been collected:

$$\mathbf{y} = (0.8,\ 1.0,\ 0.2)^{\mathrm{T}}. \tag{5.6}$$

Therefore

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} p_1 \sin(p_2) \\ p_1 \sin(2p_2) \\ p_1 \sin(3p_2) \end{pmatrix}}_{\phi(\mathbf{p})} + \underbrace{\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}}_{\mathbf{n}} \tag{5.7}$$

where $n_1 = n(1), n_2 = n(2)$, and $n_3 = n(3)$. Since $n(t)$ is assumed to be white, the vector $\mathbf{n} = (n_1, n_2, n_3)^{\mathrm{T}}$ is a white random vector (*i.e.*, n_1,n_2 and n_3 are independent). As a consequence,

[2] except for some special cases such as when ϕ is linear and when $\pi_n(\mathbf{n})$ and $\pi_{\text{prior}}(\mathbf{p})$ are Gaussian

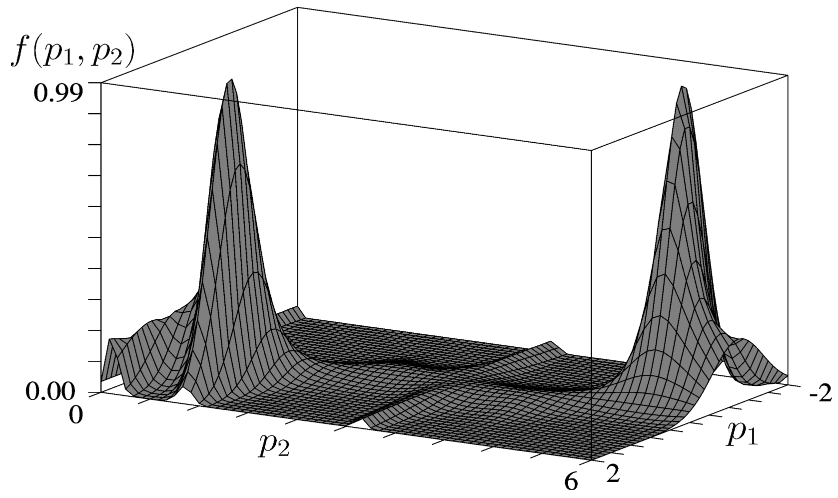

Figure 5.1: Representation of the posterior unnormalized pdf for the parameter vector $\mathbf{p}$

$$\begin{aligned} \pi_n(\mathbf{n}) &= \pi_n(n_1).\pi_n(n_2).\pi_n(n_3) && (5.8)\\ &= \frac{1}{\left(\sqrt{2\pi}\right)^3}\exp(-2n_1^2)\exp(-2n_2^2)\exp(-2n_3^2). && (5.9)\end{aligned}$$

Assume that the prior pdf for $\mathbf{p}$ is

$$\pi_{\text{prior}}(\mathbf{p}) = \frac{\text{door}_{[-2,2]}(p_1).\text{door}_{[0,6]}(p_2)}{24}, \quad (5.10)$$

where $\text{door}_{[a,b]}(p)$ is the function which takes the value 1 for $p \in [a, b]$ and zero elsewhere. This choice for $\pi_{\text{prior}}(\mathbf{p})$ can be interpreted as a prior knowledge that $\mathbf{p}$ belongs to the box $[-2, 2] \times [0, 6]$. From Equation (5.3), we get a posterior unnormalized pdf for $\mathbf{p}$:

$$f(\mathbf{p}) = \left(\prod_{k=1}^{3} \exp(-2\,(y_k - p_1 \sin(kp_2))^2)\right).\text{door}_{[-2,2]}(p_1).\text{door}_{[0,6]}(p_2). \quad (5.11)$$

This function is depicted on Figure 5.1.

The algorithm presented on Section 4 made it possible to obtain, via the solver BAYES [5], for 6 different values of α, the optimal confidence regions represented on Figure 5.2. If $\mathcal{K}_\alpha^-$ denotes the subpaving containing all dark grey boxes, and if $\mathcal{K}_\alpha^+$ denotes the subpaving containing all dark grey and white boxes, then, the interval subpaving $[\mathcal{K}_\alpha^-, \mathcal{K}_\alpha^+]$ is proved to contain $\mathbb{S}_\alpha$ (*i.e.*, $\mathbb{S}_\alpha \in [\mathcal{K}_\alpha^-, \mathcal{K}_\alpha^+]$). Note that $\mathcal{K}_0^- = \emptyset$, whereas $\mathcal{K}_1^+ = [-2, 2] \times [0, 6]$. This is consistent with the fact that $\mathbb{S}_0 = \emptyset$ and $\mathbb{S}_1 = [-2, 2] \times [0, 6]$. The computing time to get all pictures of Figure 5.2 is about 100 seconds on a Pentium 3. The interval function $[h](s)$ represented on Figure 5.3 has been computed to get Figure 5.2 and shows the relation between the threshold s and the corresponding level α. The poor quality of the enclosure, mainly due to the dependency effect occurring in Equation (5.11), could be increased at the cost of higher computing time.

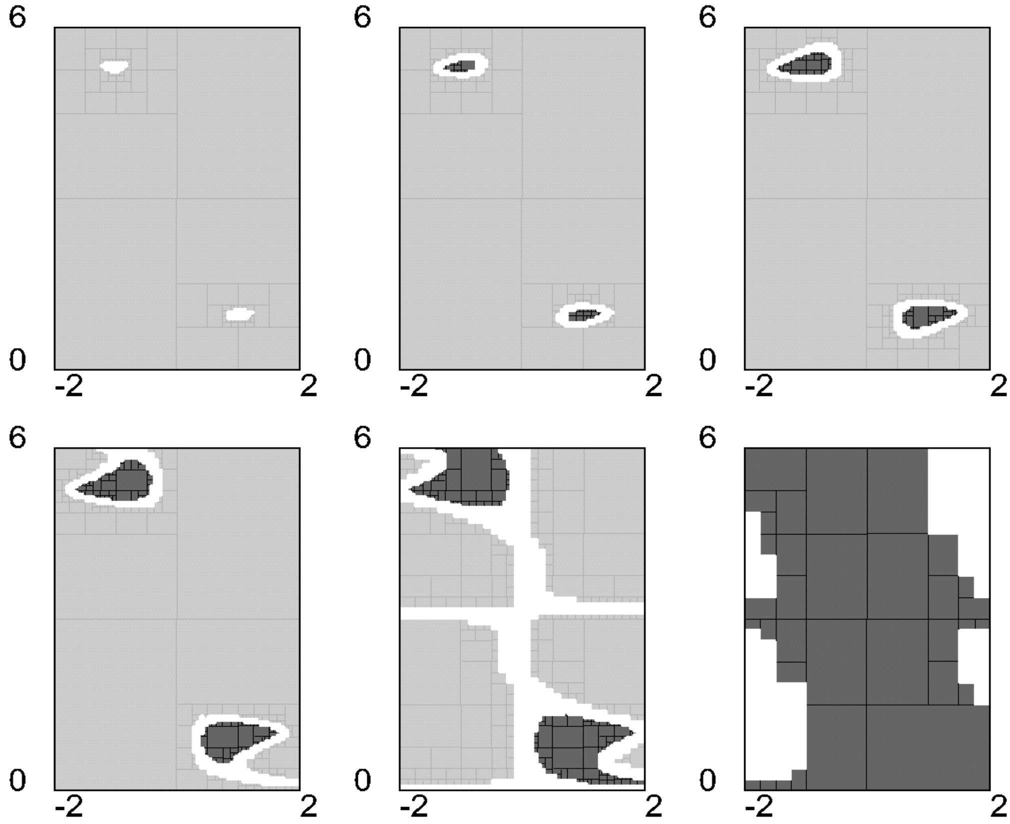

Figure 5.2: Representation of $\mathbb{S}_\alpha$ for $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$; the dark grey zone is proved to be included in $\mathbb{S}_\alpha$; the light grey zone in proved to be outside $\mathbb{S}_\alpha$; nothing is known about the white zone

6. Conclusion

In this paper, we have presented a new interval-based algorithm which made it possible to characterize the optimal confidence region of an unnormalized probability density function. To my knowledge, this problem had never been solved before in its general form.

The small application, presented in Section 5, has shown the main limitation of our algorithm: even for small problems, for two parameters and for a rough accuracy, the computing time is relatively high (here 100 seconds). When hundred of data are involved in our estimation problem, the expression of $f(\mathbf{p})$ (see Equation (5.11)) becomes much longer and the dependency effect becomes outrageous. Our algorithm would thus be unable not generate any useful results with present computers. As a consequence, this algorithm remains to be improved to be useful in practice. Inclusion functions could, for instance, be made more accurate by using the centred form instead of the natural inclusion function used here. One should also study how constraint propagation as well as interval integration techniques could be introduced in the algorithm.

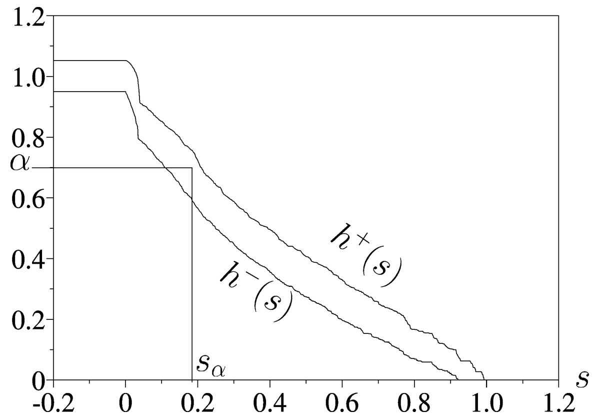

Figure 5.3: Interval function $[h](s)$ representing the relation between the threshold s and the corresponding level α

References

[1] F. Benhamou, F. Goualard, L. Granvilliers, and J. F. Puget. Revising hull and box consistency. In *Proceedings of the International Conference on Logic Programming*, pages 230–244, Las Cruces, NM, 1999.

[2] P. Eykhoff. *System Identification, Parameter and State Estimation.* John Wiley, London, 1974.

[3] E. R. Hansen. *Global Optimization using Interval Analysis.* Marcel Dekker, New York, NY, 1992.

[4] L. Jaulin. *Solution globale et garantie de problèmes ensemblistes ; application à l'estimation non linéaire et à la commande robuste.* PhD dissertation, Université Paris-Sud, Orsay, France, 1994. Available at: http://www.istia.univ-angers.fr/~jaulin/thesejaulin.zip.

[5] L. Jaulin. *BAYES, a Windows solver to characterize optimal confidence regions, available at* `http://www.istia.univ-angers.fr/~jaulin/bayes.zip`. E3I2, ENSIETA, 2004.

[6] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics.* Springer-Verlag, London, 2001.

[7] R. E. Moore. *Methods and Applications of Interval Analysis.* SIAM, Philadelphia, PA, 1979.

[8] P. Nivlet. *Prise en compte des incertitudes dans l'interprétation réservoir des données géophysiques ; application à la segmentation.* PhD dissertation, Institut Français du Pétrole, France, 2001.

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