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

M. Kieffer

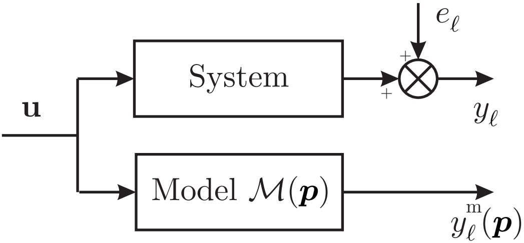
L2S, CNRS - Supélec - Univ Paris-Sud

9 july 2017

Outline

1 Introduction
2 Estimating parameter and uncertainty
- Classical approaches
- Approaches proposed by Campi et al.
3 LSCR
4 SPS
5 Guaranteed characterization via interval analysis
6 Example (LSCR)
7 Contractors
8 Example (SPS)
- Low-dimensional model
- Higher-dimensional model
9 Conclusions and perspectives

Introduction



- Parameter identification
  - estimate value of parameter vector $\mathbf{p}$
  - considering some model structure $\mathscr{M}(\cdot)$
  - from noisy data vector $\mathbf{y}$

Introduction

- Usually via minimization of cost function, 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 vector of model outputs
  - $\|\cdot\|_2$ is a (possibly weighted) ℓ_2 norm.

- Then

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

Difficulties

- Parameters of model may not be identifiable uniquely
  $\hookrightarrow$ different values of $\hat{\mathbf{p}}$ may yield the same $\mathbf{y}_{\mathrm{m}}(\hat{\mathbf{p}})$

- Numerical algorithm to compute $\hat{\mathbf{p}}$ may get trapped at local minimizer

- Even if single $\hat{\mathbf{p}}$ is obtained and if $\mathbf{y} \simeq \mathbf{y}_{\mathrm{m}}(\hat{\mathbf{p}})$, $\hat{\mathbf{p}}$ cannot be considered as final answer to the estimation problem
  $\hookrightarrow$ quality tag is missing.

$$\hat{p}_i = 1.2345 \pm 10^{-4} \text{ is quite different of } \hat{p}_i = 1.2345 \pm 10^{3}.$$

Outline

1. Introduction
2. Estimating parameter and uncertainty
   - Classical approaches
   - Approaches proposed by Campi et al.
3. LSCR
4. SPS
5. Guaranteed characterization via interval analysis
6. Example (LSCR)
7. Contractors
8. Example (SPS)
   - Low-dimensional model
   - Higher-dimensional model
9. Conclusions and perspectives

Classical approaches

Based on

- Level-set [11].
- Monte-Carlo techniques [11].
- Evaluation of the density of the estimator [8].
- Bounded-error estimation [9].

Characterization of parameter uncertainty via previous approaches relies on hypotheses on noise corrupting data

- difficult to check from residuals $\mathbf{y} - \mathbf{y}_{\mathrm{m}}(\hat{\mathbf{p}})$ when n_{y} is large,
- impossible to check when only few data points.

LSCR and SPS

Campi *et al.* [1, 3, 2] propose two new approaches named LSCR and SPS

- *exact characterization* of parameter uncertainty
- in *non-asymptotic* conditions.

Hypotheses

1. System generating data must belong to model set (true value $\mathbf{p}^*$ should be meaningful)
2. Noise samples must be independently distributed with distributions symmetric with respect to zero.

Outline

1. Introduction
2. Estimating parameter and uncertainty
   - Classical approaches
   - Approaches proposed by Campi et al.
3. LSCR
4. SPS
5. Guaranteed characterization via interval analysis
6. Example (LSCR)
7. Contractors
8. Example (SPS)
   - Low-dimensional model
   - Higher-dimensional model
9. Conclusions and perspectives

LSCR

Introduction - main idea

LSCR [1]: *leave-out sign-dominant correlated regions*

- Independent estimates of the correlation of the prediction error

$$\varepsilon_t(\mathbf{p}) = y_t - y_t^{\mathrm{m}}(\mathbf{p})$$

  should have random signs.
- Leave out subset of parameter space where sign does not appear random (*i.e.* is sign dominant)

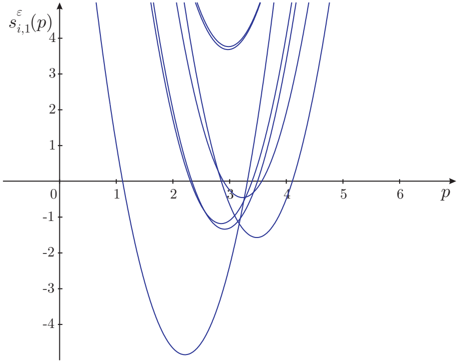
Defines, without any approximation,

region $\mathbf{\Theta}$ to which $\mathbf{p}^*$ belongs with specified probability.

LSCR

Example

Model $y_t^{\mathrm{m}}(p) = p$, with 8 noisy data generated with $p^* = 3$.



7 different empirical correlations

LSCR I

Description

Consider prediction error

$$\varepsilon_t(\mathbf{p}) = y_t - y_t^{\mathrm{m}}(\mathbf{p})$$

such that $\varepsilon_t(\mathbf{p}^*)$ is realization of noise corrupting data at time t.

1. Select two integers $r \geqslant 0$ and $q \geqslant 0$.
2. For $t = 1 + r, \ldots, k + r = n$, compute

$$c_{t-r,r}^{\varepsilon}(\mathbf{p}) = \varepsilon_{t-r}(\mathbf{p})\,\varepsilon_t(\mathbf{p})\,. \tag{3}$$

LSCR II

Description

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 $\mathbb{I}_i \subset \mathbb{I}$, set of indexes. Collection $\mathbb{G}$ of 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 $\mathbf{\Theta}^{\varepsilon}_{r,q}$ such that at least q of functions $s^{\varepsilon}_{i,r}(\mathbf{p})$ are larger than 0 and at least q are smaller than 0.

LSCR

Description

Exemple of $\mathbb{G}$ st $\forall \mathbb{I}_i \in \mathbb{G}$,$\forall \mathbb{I}_i \in \mathbb{G}$ one has $(\mathbb{I}_i \cup \mathbb{I}_j) - (\mathbb{I}_i \cap \mathbb{I}_j) \in \mathbb{G}$

	1	2	3	4	5	6	7
$\mathbb{I}_1$	•	•		•	•		
$\mathbb{I}_2$	•		•	•		•	
$\mathbb{I}_3$		•	•		•	•	
$\mathbb{I}_4$	•	•				•	•
$\mathbb{I}_5$	•		•		•		•
$\mathbb{I}_6$		•	•	•			•
$\mathbb{I}_7$				•	•	•	•
$\mathbb{I}_8$							

$$s^{\varepsilon}_{1,r=1}(\mathrm{p}) = \varepsilon_1(\mathrm{p})\varepsilon_2(\mathrm{p}) + \varepsilon_2(\mathrm{p})\varepsilon_3(\mathrm{p}) + \varepsilon_4(\mathrm{p})\varepsilon_5(\mathrm{p}) + \varepsilon_5(\mathrm{p})\varepsilon_6(\mathrm{p})$$
$$s^{\varepsilon}_{2,r=1}(\mathrm{p}) = \varepsilon_1(\mathrm{p})\varepsilon_2(\mathrm{p}) + \varepsilon_3(\mathrm{p})\varepsilon_4(\mathrm{p}) + \varepsilon_4(\mathrm{p})\varepsilon_5(\mathrm{p}) + \varepsilon_6(\mathrm{p})\varepsilon_7(\mathrm{p})$$
$$s^{\varepsilon}_{3,r=1}(\mathrm{p}) = \ldots$$

LSCR

Properties

The set $\boldsymbol{\Theta}^{\varepsilon}_{r,q}$ is such that [1]

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

Shape and size of $\boldsymbol{\Theta}^{\varepsilon}_{r,q}$ depend on

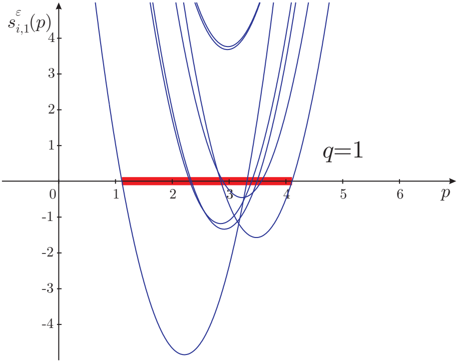
- values given to q and r
- group $\mathbb{G}$ and its number of elements m.

A procedure for generating $\mathbb{G}$ of appropriate size suggested in [4].

LSCR

Example (continued)

Model $y_t^m(p) = p$, with 8 noisy data generated with $p^* = 3$.



7 empirical correlations, and 71% confidence region

LSCR

More formal definition

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

$$\mathbf{\Theta}^{\varepsilon}_{r,q} = \mathbf{\Theta}^{\varepsilon,-}_{r,q} \cap \mathbf{\Theta}^{\varepsilon,+}_{r,q}, \tag{5}$$

with

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

$$\mathbf{\Theta}^{\varepsilon,+}_{r,q} = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i^{\varepsilon,+}(\mathbf{p}) \geqslant q \right\}, \tag{7}$$

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

LSCR

More formal definition

Moreover

$$\tau_i^{\varepsilon,-}(\mathbf{p}) = \begin{cases} 1 & \text{if } -s_{i,r}^{\varepsilon}(\mathbf{p}) \geqslant 0, \\ 0 & \text{else,} \end{cases} \tag{8}$$

and

$$\tau_i^{\varepsilon,+}(\mathbf{p}) = \begin{cases} 1 & \text{if } s_{i,r}^{\varepsilon}(\mathbf{p}) \geqslant 0, \\ 0 & \text{else.} \end{cases} \tag{9}$$

$\mathbf{\Theta}_{r,q}^{\varepsilon,-}$ 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 $\mathbf{\Theta}_{r,q}^{\varepsilon,+}$ 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.

Outline

1 Introduction
2 Estimating parameter and uncertainty
- Classical approaches
- Approaches proposed by Campi et al.
3 LSCR
4 SPS
5 Guaranteed characterization via interval analysis
6 Example (LSCR)
7 Contractors
8 Example (SPS)
- Low-dimensional model
- Higher-dimensional model
9 Conclusions and perspectives

SPS

Introduction

SPS [2]: *sign-perturbed sums*.

SPS is designed for linear regression, where

$$y_t = \varphi_t^\mathsf{T} \mathbf{p}^* + w_t, t = 1, \ldots, n, \tag{10}$$

with φ_t known regression vector.

SPS computes an exact confidence region for $\mathbf{p}^*$ around least-squares estimate $\hat{\mathbf{p}}$, which is solution to *normal equations*

$$\sum_{t=1}^{n} \varphi_t \left(y_t - \varphi_t^\mathsf{T} \hat{\mathbf{p}} \right) = \mathbf{0}. \tag{11}$$

SPS

Description

For a generic **p**, define

$$s_0(\mathbf{p}) = \sum_{t=1}^{n} \varphi_t \left(y_t - \varphi_t^T \mathbf{p} \right), \tag{12}$$

and the sign-perturbed sums

$$s_i(\mathbf{p}) = \sum_{t=1}^{n} \alpha_{i,t}\varphi_t \left(y_t - \varphi_t^T \mathbf{p} \right), \tag{13}$$

where $i = 1, \ldots, m-1$ and $\alpha_{i,t} = \pm 1$ with equal probability, and

$$z_i(\mathbf{p}) = \|s_i(\mathbf{p})\|_2^2, i = 0, \ldots, m-1. \tag{14}$$

SPS

Description

Confidence region $\mathbf{\Sigma}_q$ is set of all $\mathbf{p}$ such that $z_0(\mathbf{p})$ is *not* among the q largest values of $(z_i(\mathbf{p}))_{i=0}^{m-1}$.

In [2], it is shown that $\mathbf{p}^* \in \mathbf{\Sigma}_q$ with exact probability $1 - q/m$.

SPS

More formal definition

$\mathbf{\Sigma}_q$ may be defined more formally as

$$\mathbf{\Sigma}_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m-1} \tau_i(\mathbf{p}) \geqslant q \right\} \tag{15}$$

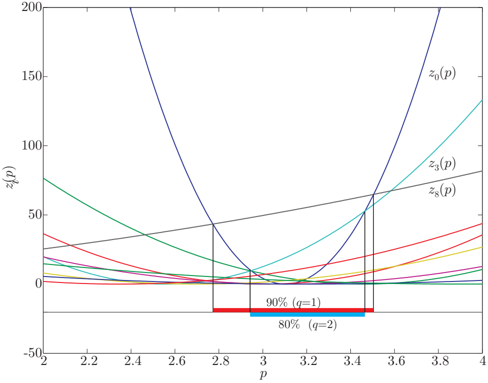
where

$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } z_i(\mathbf{p}) - z_0(\mathbf{p}) > 0, \\ 0 & \text{else.} \end{cases} \tag{16}$$

SPS

Illustration

Model $y_t^m(p) = p$, with 20 noisy data generated for $p^* = 3$. We choose $m = 10$.



Outline

1 Introduction
2 Estimating parameter and uncertainty
- Classical approaches
- Approaches proposed by Campi et al.

3 LSCR
4 SPS
5 Guaranteed characterization via interval analysis
6 Example (LSCR)
7 Contractors
8 Example (SPS)
- Low-dimensional model
- Higher-dimensional model

9 Conclusions and perspectives

Guaranteed characterization

In LSCR (and SPS), one has to characterize

$$\boldsymbol{\Psi}_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q \right\}, \tag{17}$$

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

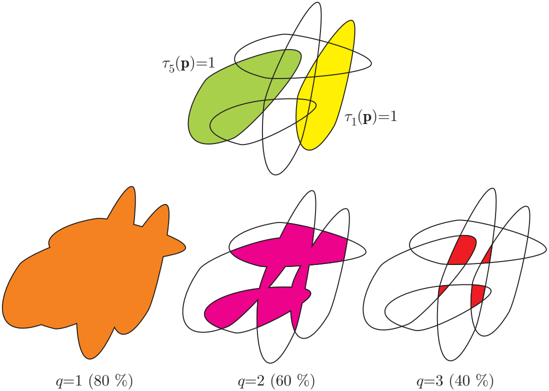
$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } f_i(\mathbf{p}) \geqslant 0, \\ 0 & \text{else,} \end{cases} \tag{18}$$

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

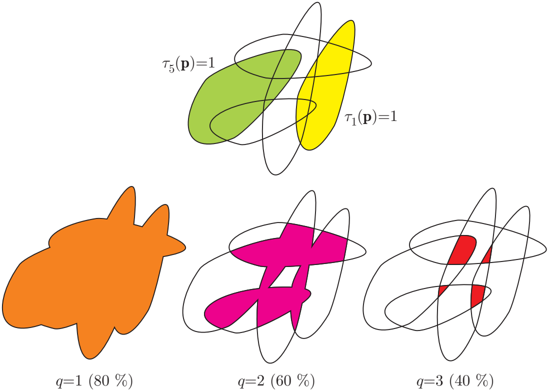
Guaranteed characterization

In LSCR (and SPS), one has to characterize

$$\boldsymbol{\Psi}_q = \left\{ \mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q \right\}, \tag{19}$$



Guaranteed characterization



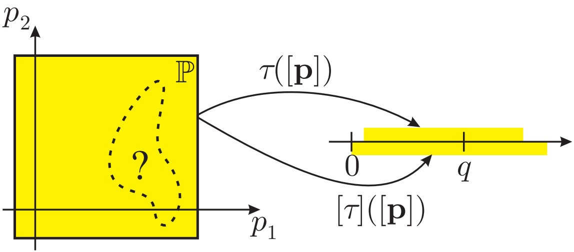
Characterization

- approximate using gridding in [1, 3, 2].
- guaranteed using interval analysis here.

SIVIA

To characterize $\mathbf{\Psi}_q = \{\mathbf{p} \in \mathbb{P}$ such that $\sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q\}$, one uses SIVIA and an inclusion function [10, ?] $[\tau]([\mathbf{p}])$ of

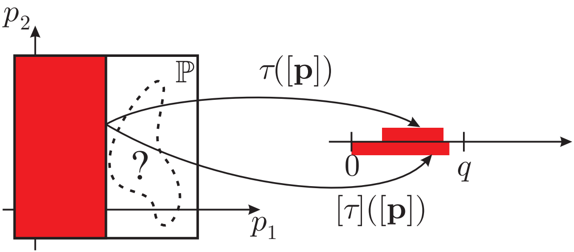
$$\tau(\mathbf{p}) = \sum_{i=1}^{m} \tau_i(\mathbf{p}).$$



SIVIA

To characterize $\mathbf{\Psi}_q = \{\mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q\}$, one uses an inclusion function $[\tau]([\mathbf{p}])$ of

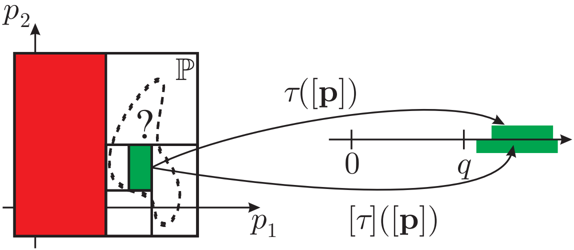
$$\tau(\mathbf{p}) = \sum_{i=1}^{m} \tau_i(\mathbf{p}).$$



SIVIA

To characterize $\boldsymbol{\Psi}_q = \{\mathbf{p} \in \mathbb{P} \text{ such that } \sum_{i=1}^{m} \tau_i(\mathbf{p}) \geqslant q\}$, one uses an inclusion function $[\tau]([\mathbf{p}])$ of

$$\tau(\mathbf{p}) = \sum_{i=1}^{m} \tau_i(\mathbf{p}).$$

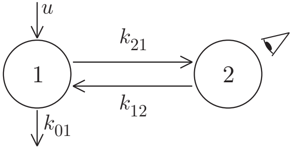


Outline

1. Introduction
2. Estimating parameter and uncertainty
   - Classical approaches
   - Approaches proposed by Campi et al.
3. LSCR
4. SPS
5. Guaranteed characterization via interval analysis
6. **Example (LSCR)**
7. Contractors
8. Example (SPS)
   - Low-dimensional model
   - Higher-dimensional model
9. Conclusions and perspectives

System and Model

Consider the two-compartment model



System and Model

System output obtained from

$$y_t = \alpha(\mathbf{p}^*)(\exp(\lambda_1(\mathbf{p}^*)t) - \exp(\lambda_2(\mathbf{p}^*)t)) + w_t, \tag{20}$$

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

$$\alpha(\mathbf{p}) = \frac{k_{21}}{\sqrt{(k_{01} - k_{12} + k_{21})^2 + 4k_{12}k_{21}}},$$

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

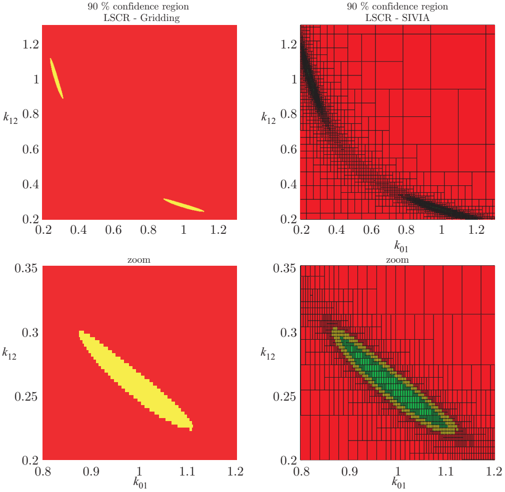
and w_t's are realizations of iid $\mathcal{N}(0, \sigma^2)$ variables, for $t = 0, T, \ldots, (n-1)T$.

System and Model

Data generated with $\mathbf{p}^* = (1, 0.25, 0.5)^\mathsf{T}$, $\sigma^2 = 10^{-4}$.
Sampling period is $T = 0.02$ s and $n = 64$.
Only k_{01} et k_{12} are estimated, value k_{21}^* of k_{21} assumed known.
Measurement noise is additive, LSCR method applies directly.

Confidence region obtained by LSCR

$\mathbb{P} = [0,5] \times [0,5]$ and $\varepsilon = 0.0025$.



Outline

1. Introduction
2. Estimating parameter and uncertainty
   - Classical approaches
   - Approaches proposed by Campi et al.
3. LSCR
4. SPS
5. Guaranteed characterization via interval analysis
6. Example (LSCR)
7. Contractors
8. Example (SPS)
   - Low-dimensional model
   - Higher-dimensional model
9. Conclusions and perspectives

Contractors

Introduction

Contractor $\mathscr{C}_{\mathbf{f},\mathbb{Y}}$ associated with generic set-inversion problem

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

takes $[\mathbf{x}]$ as input and returns

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

such that

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

so no part of $\mathbb{X}$ in $[\mathbf{x}]$ is lost.

Contractors

Examples

Various types of contractors

- by interval constraint propagation,
- by parallel linearization,
- the Newton contractor,
- the Krawczyk contractor, *etc.*

Contractors

With LSCR and SPS

$$\boldsymbol{\Psi}_q = \mathbb{P} \cap \tau^{-1}([q, m]), \tag{24}$$

The τs are not differentiable and forbid use of classic contractors.

$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } f_i(\mathbf{p}) \geqslant 0, \\ 0 & \text{else,} \end{cases} \tag{25}$$

New proposed contractor assumes f_is are differentiable.

1. build set of m possibly overlapping subboxes of $[\mathbf{p}]$, trying to remove all values of $\mathbf{p}$ from $[\mathbf{p}]$ such that $f_i(\mathbf{p}) < 0$, $i = 1, \ldots, m$.
2. compute union of all non-empty intersections of at least q of these boxes.

Contractors

With LSCR and SPS

$$\mathbf{\Psi}_q = \mathbb{P} \cap \tau^{-1}([q, m]), \tag{24}$$

The τs are not differentiable and forbid use of classic contractors.

$$\tau_i(\mathbf{p}) = \begin{cases} 1 & \text{if } f_i(\mathbf{p}) \geqslant 0, \\ 0 & \text{else}, \end{cases} \tag{25}$$

New proposed contractor assumes f_is are differentiable.

1. build set of m possibly overlapping subboxes of $[\mathbf{p}]$, trying to remove all values of $\mathbf{p}$ from $[\mathbf{p}]$ such that $f_i(\mathbf{p}) < 0$, $i = 1, \ldots, m$.
2. compute union of all non-empty intersections of at least q of these boxes.

Box contraction using the f_i's, suitable for LSCR and SPS

First step uses centered inclusion function of f_i. For some $\mathbf{m} \in [\mathbf{p}]$, may be written as

$$[f_{i,\mathrm{c}}]([\mathbf{p}]) = f_i(\mathbf{m}) + ([\mathbf{p}] - \mathbf{m})^{\mathrm{T}} [\mathbf{g}_i]([\mathbf{p}]) \tag{26}$$

$$= f_i(\mathbf{m}) + \sum_{j=1}^{n_{\mathrm{p}}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]), \tag{27}$$

where $\mathbf{g}_i$ is gradient of f_i.

Box contraction using the f_i's, suitable for LSCR and SPS

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

$$\left[p'_{i,k}\right] = [p_k] \cap \Bigg(\Bigg(([f_{i,\mathrm{c}}]([\mathbf{p}]) \cap [0,\infty[) - f_i(\mathbf{m}) \\ - \sum_{j=1, j\neq k}^{n_{\mathrm{p}}} ([p_j] - m_j)[g_{i,j}]([\mathbf{p}]) \Bigg) / [g_{i,k}]([\mathbf{p}]) + m_k \Bigg). \tag{28}$$

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

$$\left[p'_{i,k}\right] = [p_k]. \tag{29}$$

Box contraction using the f_i's, suitable for LSCR and SPS

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

$$\mathscr{L} = \left\{ \mathscr{C}_{f_1,[0,\infty[}\left([\mathbf{p}]\right), \ldots, \mathscr{C}_{f_m,[0,\infty[}\left([\mathbf{p}]\right) \right\} \tag{30}$$

$$= \left\{ \left[\mathbf{p}_1'\right], \ldots, \left[\mathbf{p}_m'\right] \right\}. \tag{31}$$

Here, $[\mathbf{p}_i'] = \emptyset$ indicates that there is no $\mathbf{p} \in [\mathbf{p}]$ such that $f_i\left(\mathbf{p}\right) \geqslant 0$.

Our aim is to evaluate a subbox $[\mathbf{p}']$ of $[\mathbf{p}]$ such that $\boldsymbol{\Psi}_q \cap [\mathbf{p}'] = \boldsymbol{\Psi}_q \cap [\mathbf{p}]$.

Box contraction suitable for SPS only

Main idea

Takes advantage of $\mathbf{s}_i(\mathbf{p})$, $i = 0, \ldots, m$ affine in $\mathbf{p}$ to

- reduce number of occurrences of $\mathbf{p}$ in $\mathbf{s}_i(\mathbf{p})$,
- reduce pessimism of corresponding inclusion functions.

Box contraction suitable for SPS only

One may rewrite

$$s_0(\mathbf{p}) = \sum_{t=1}^{n} \varphi_t \left(y_t - \varphi_t^{\mathsf{T}} \mathbf{p} \right),$$

as

$$\begin{aligned} s_0(\mathbf{p}) &= \sum_{t=1}^{n} y_t \varphi_t - \left(\sum_{t=1}^{n} \varphi_t \varphi_t^{\mathsf{T}} \right) \mathbf{p} \\ &= \mathbf{b}_0 - \mathbf{A}_0 \mathbf{p} \end{aligned}$$

with $\mathbf{b}_0 = \sum_{t=1}^{n} y_t \varphi_t$ and $\mathbf{A}_0 = \left(\sum_{t=1}^{n} \varphi_t \varphi_t^{\mathsf{T}} \right)$.

Box contraction suitable for SPS only

Similarly, one may write

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

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

Box contraction suitable for SPS only

Using

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

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

one gets

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

The matrices $\mathbf{A}_i^2 - \mathbf{A}_0^2$ are symmetric

$$\mathbf{A}_i^2 - \mathbf{A}_0^2 = \mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U}.$$

Using the change of variables $\pi = \mathbf{U}\mathbf{p}$, $z_i(\mathbf{p}) - z_0(\mathbf{p})$ becomes

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \pi^\mathsf{T}\mathbf{D}\pi - 2\beta^\mathsf{T}\pi + \gamma,$$

with $\beta^\mathsf{T} = (\mathbf{b}_i^\mathsf{T}\mathbf{A}_i - \mathbf{b}_0^\mathsf{T}\mathbf{A}_0)\mathbf{U}^\mathsf{T}$ and $\gamma = \mathbf{b}_i^\mathsf{T}\mathbf{b}_i - \mathbf{b}_0^\mathsf{T}\mathbf{b}_0$.

Box contraction suitable for SPS only

Using

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

one gets

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

The matrices $\mathbf{A}_i^2 - \mathbf{A}_0^2$ are symmetric

$$\mathbf{A}_i^2 - \mathbf{A}_0^2 = \mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U}.$$

Using the change of variables $\pi = \mathbf{U}\mathbf{p}$, $z_i(\mathbf{p}) - z_0(\mathbf{p})$ becomes

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \pi^\mathsf{T}\mathbf{D}\pi - 2\beta^\mathsf{T}\pi + \gamma,$$

with $\beta^\mathsf{T} = \left(\mathbf{b}_i^\mathsf{T}\mathbf{A}_i - \mathbf{b}_0^\mathsf{T}\mathbf{A}_0\right)\mathbf{U}^\mathsf{T}$ and $\gamma = \mathbf{b}_i^\mathsf{T}\mathbf{b}_i - \mathbf{b}_0^\mathsf{T}\mathbf{b}_0$.

Box contraction suitable for SPS only

One then obtains

$$z_i(\mathbf{p}) - z_0(\mathbf{p}) = \sum_{i=1}^{n_\mathrm{p}} d_i \left(\pi_i - \frac{\beta_i}{d_i}\right)^2 + \gamma - \sum_{i=1}^{n_\mathrm{p}} \frac{\beta_i^2}{d_i}.$$

If $\mathbf{p} \in [\underline{\mathbf{p}}, \overline{\mathbf{p}}]$, one is able to get $\pi \in [\underline{\pi}, \overline{\pi}] = \mathbf{U}[\underline{\mathbf{p}}, \overline{\mathbf{p}}]$.
Whenever $d_i \neq 0$, a contractor for $[\pi_i]$ is then obtained

$$[\pi_i'] = [\pi_i] \cap \left(\pm \sqrt{\left(([z_i - z_0]([\mathbf{p}]) \cap [0, \infty[) - \sum_{j=1, j\neq i}^{n_\mathrm{p}} d_j \left(\pi_j - \frac{\beta_j}{d_j}\right)^2 - \gamma + \sum_{i=1}^{n_\mathrm{p}} \frac{\beta_i^2}{d_i} \right) / d_i \cdots}\right.$$

If $d_i = 0$, $[\pi_i]$ is left unchanged. Then, a contractor for $[\mathbf{p}]$ is obtained as

$$[\mathbf{p}'] = [\mathbf{p}] \cap \left(\mathbf{U}^\mathrm{T} [\pi'] \right).$$

Building a q-relaxed intersection

Second step: contractor builds a box $[\mathbf{p}']$ enclosing the q-relaxed intersection $\mathscr{P}$ [7, 5, 6] of the boxes in $\mathscr{L} = \{[\mathbf{p}'_1], \ldots, [\mathbf{p}'_m]\}$

$$\mathscr{P} = \bigcap_{j\in\{1,\ldots,m\}}^{q} [\mathbf{p}'_j]\,. \tag{32}$$

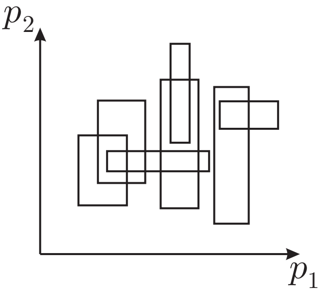
$$= \bigcup_{\substack{J\subset[1,\ldots,m]\\ \mathrm{card}\,(J)\geqslant q}} \bigcap_{j\in J} [\mathbf{p}'_j]\,, \tag{33}$$

and satisfying

$$\mathscr{P} \subset [\mathbf{p}'] \subset [\mathbf{p}]\,. \tag{34}$$

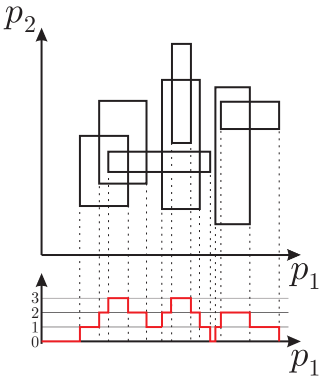
Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \ldots, [p_m]\}$ of m intervals.



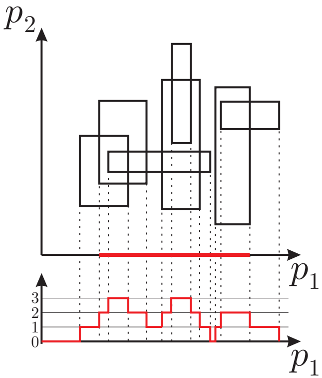
Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \ldots, [p_m]\}$ of m intervals.



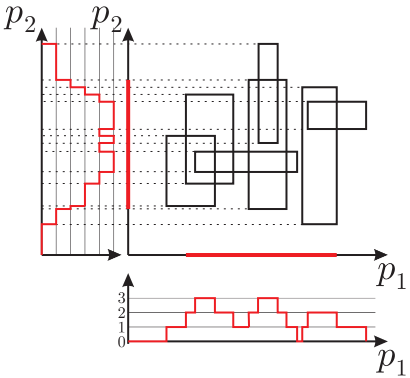
Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \ldots, [p_m]\}$ of m intervals.



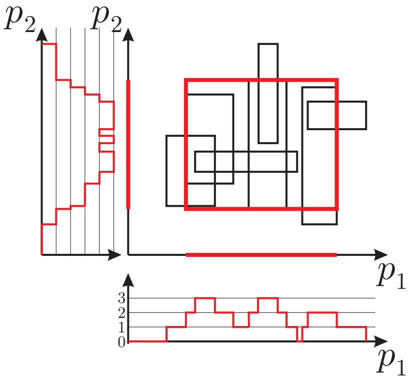
Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \ldots, [p_m]\}$ of m intervals.



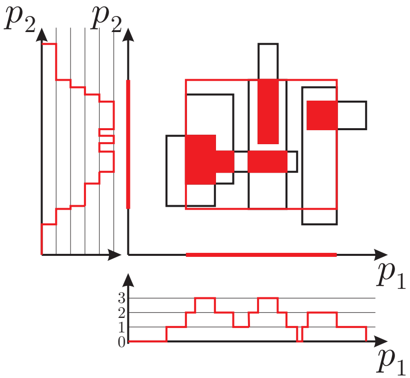
Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \ldots, [p_m]\}$ of m intervals.



Evaluating the q-relaxed intersection

Consider a list $\mathscr{L} = \{[p_1], \dots, [p_m]\}$ of m intervals.



Outline

1. Introduction
2. Estimating parameter and uncertainty
   - Classical approaches
   - Approaches proposed by Campi et al.
3. LSCR
4. SPS
5. Guaranteed characterization via interval analysis
6. Example (LSCR)
7. Contractors
8. Example (SPS)
   - Low-dimensional model
   - Higher-dimensional model
9. Conclusions and perspectives

System and model

Consider system such that

$$u_t = \alpha u_{t-1} + v_t, \tag{35}$$

$$y_t = a^* u_t + b^* u_{t-1} + w_t, \tag{36}$$

with $\alpha = 0.5$ and $u_0 = 0$.

For $t = 1, \ldots, n$, v_t and w_t are iid $\mathcal{N}\left(0, \sigma^2\right)$.

Take as a model

$$y_t^{\mathrm{m}}(\mathbf{p}) = a u_t + b u_{t-1}, \tag{37}$$

which is linear in $\mathbf{p} = (a, b)^{\mathrm{T}}$.

With SPS

In linear regression form, one gets

$$y_t = \varphi_t^{\mathsf{T}} \mathbf{p}^* + w_t \tag{38}$$

with $\varphi_t = (u_t, u_{t-1})^{\mathsf{T}}$ and $\mathbf{p}^* = (a^*, b^*)^{\mathsf{T}}$.

Characterization of $\boldsymbol{\Sigma}_q$ addressed using SIVIA.

Inclusion functions for τ_i's are introduced

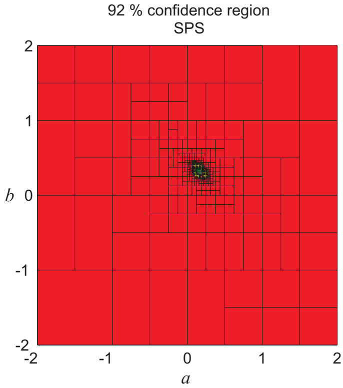
$$[\tau_i]([\mathbf{p}]) = \begin{cases} 1 & \text{if } \inf([f_i]([\mathbf{p}])) \geqslant 0, \\ 0 & \text{if } \sup([f_i]([\mathbf{p}])) < 0, \\ [0,1] & \text{else,} \end{cases} \tag{39}$$

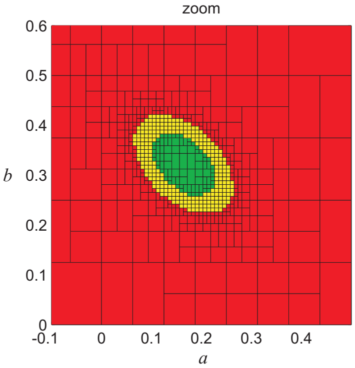
where

$$[f_i]([\mathbf{p}]) = [z_i - z_0]([\mathbf{p}]). \tag{40}$$

Natural inclusion function

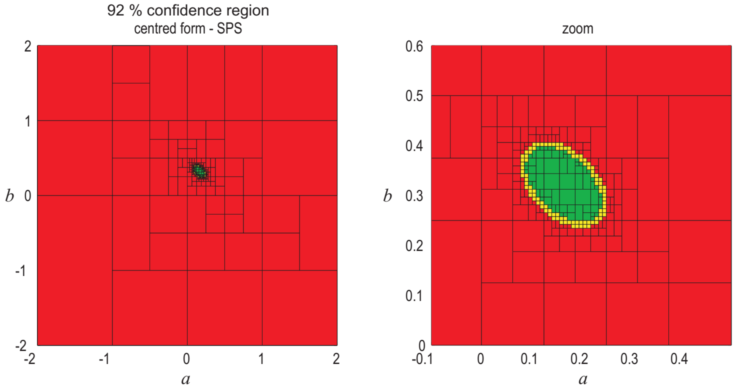
Paving of the search set $\mathbb{P} = [-2,2] \times [-2,2]$ obtained when $a = 0.2$, $b = 0.3$, $\sigma^2 = 0.25$, $n = 256$, $m = 255$, and $q = 20$.



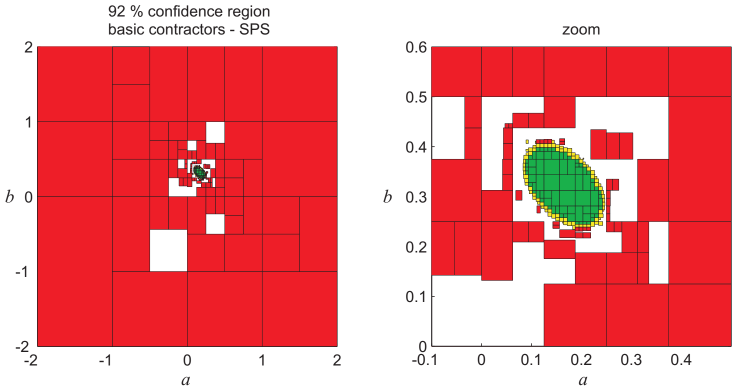


Centred forms

Paving of the search set $\mathbb{P} = [-2,2] \times [-2,2]$ obtained when $a = 0.2$, $b = 0.3$, $\sigma^2 = 0.25$, $n = 256$, $m = 255$, and $q = 20$.

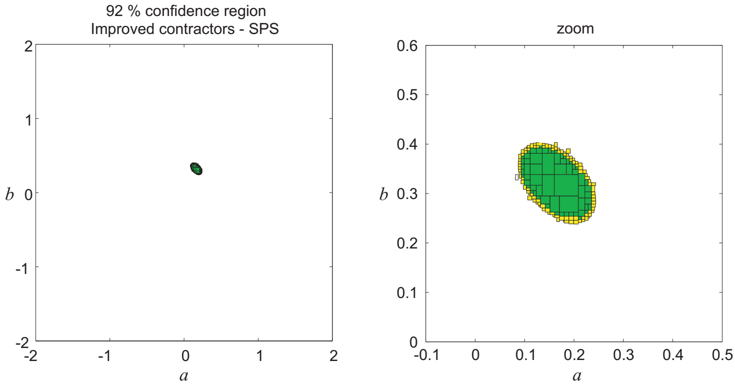


Using the LSCR/SPS contractor



Subpavings of the search space obtained using SPS in the linear case with the LSCR/SPS contractor

Using the SPS contractor



Subpavings of the search space obtained using SPS in the linear case with the SPS contractor

System and Model

Consider the system

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

with the FIR model

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

where $\mathbf{p} = (a_0, \ldots, a_{n_{\mathrm{a}}-1})^{\mathrm{T}}$ and $u_n = 0$ for $n \leqslant 0$. For $t = 1, \ldots, n$, the w_ts are iid noise samples.
In linear regression form, one has

$$y_t = \varphi_t^{\mathrm{T}} \mathbf{p}^* + w_t$$

with $\varphi_t = (u_t, \ldots, u_{t-n_{\mathrm{a}}+1})^{\mathrm{T}}$ and $\mathbf{p}^* = \left(a_0^*, \ldots, a_{n_{\mathrm{a}}-1}^*\right)^{\mathrm{T}}$.

Inclusion function

When the dimension of $\mathbf{p}$ is small, $\mathbf{\Sigma}_q$ may be characterized using SIVIA and inclusion functions for τ_i

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

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

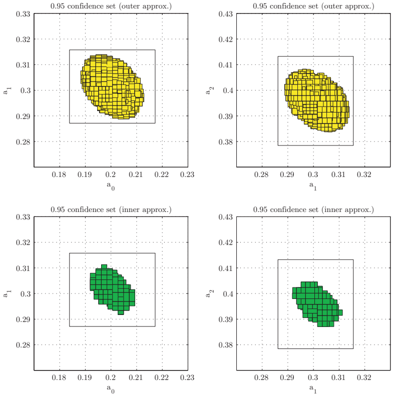
Simulation conditions

Data are generated for $a_0^* = 0.2$, $a_1^* = 0.3$, and $a_2^* = 0.4$ considering:

1. a filtered Gaussian input $u_t = \alpha u_{t-1} + v_t$, with $\alpha = 0.2$ and $v_t \sim \mathcal{N}(0, 0.65)$
2. a random iid sequence of ± 1 (D-optimal input when input has to remain in $[-1, 1]$.

w_t zero-mean Laplacian with standard deviation σ_w tuned to get a signal-to-noise ratio (SNR) of 15 dB.
We choose $n = 1024$, $m = 255$, and $q = 13$ (95 % confidence region), $\varepsilon = 2.5 \times 10^{-3}$.

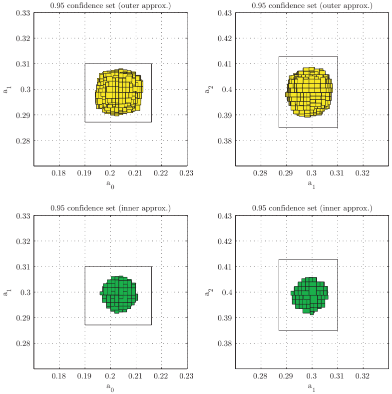
Results



Projections of the obtained outer and inner-approximations

Gaussian input

Results



Projections of the obtained outer and inner-approximations

D-optimal input

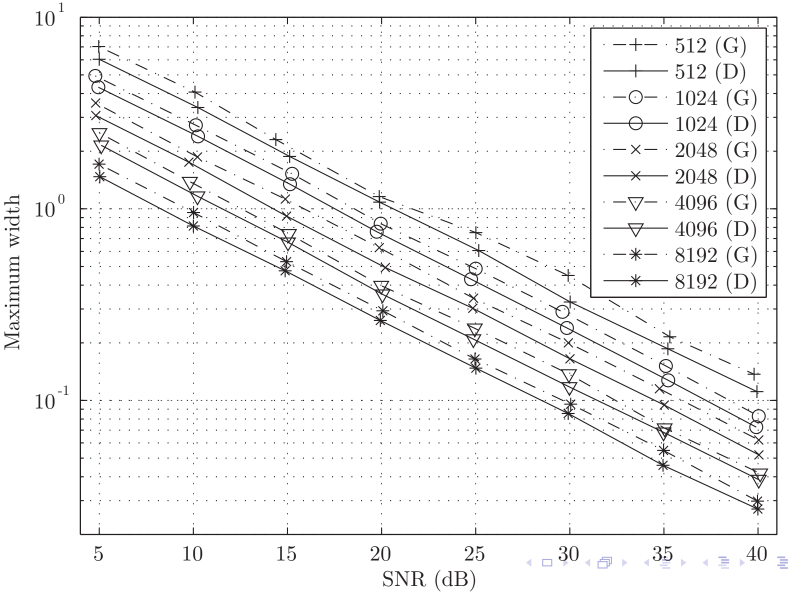
Simulation conditions

FIR models with $n_a = 20$ random parameters in $[-2,2]^{n_a}$ are generated, then

- $n = 512$, 1024, 2048, 4096, and 8192 noise-free data points are generated
- white Laplacian noise is added to the data.

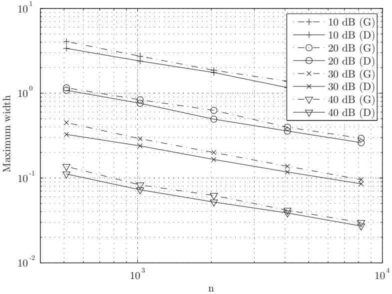
Standard deviation of noise set up to get an SNR of 5 dB to 40 dB.
We choose $n = 1024$, $m = 255$, and $q = 13$ (95 % confidence. Only outer approximations may be obtained.
Initial search box $\mathbb{P} = \left[-10^4, 10^4\right]^{20}$.

Results



10^1
10^0
10^{-1}
Maximum width
5
10
15
20
25
30
35
40
SNR (dB)
512 (G)
512 (D)
1024 (G)
1024 (D)
2048 (G)
2048 (D)
4096 (G)
4096 (D)
8192 (G)
8192 (D)

Results



Maximum width as a function of n

Slope is about $-1/2$, consistent with ML estimation with additive Gaussian noise

Outline

1 Introduction
2 Estimating parameter and uncertainty
- Classical approaches
- Approaches proposed by Campi et al.
3 LSCR
4 SPS
5 Guaranteed characterization via interval analysis
6 Example (LSCR)
7 Contractors
8 Example (SPS)
- Low-dimensional model
- Higher-dimensional model
9 Conclusions and perspectives

Conclusions and perspectives

- Interval analysis provides guaranteed outer- and inner-approximations of non-asymptotic confidence regions defined by LSCR and SPS.
- Illustrations provided for FIR and non-linear models.
- Accurate inclusion functions are particularly difficult to obtain for the functions involved in SPS,
- Symbolic manipulations of the involved expression to reduce the number of occurrences of the parameters are particularly useful to
  - improve the efficiency of SIVIA
  - to design better contractors

Code available at `http://www.l2s.supelec.fr/perso/kieffer-0`

References I

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.

References II

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.
Guaranteed robust nonlinear minimax estimation.
IEEE Transaction on Automatic Control, 47(11):1857–1864, 2002.

References III

S. M. Kay.
Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.
Prentice Hall, 1993.

M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter, editors.
Bounding Approaches to System Identification.
Plenum Press, New York, NY, 1996.

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

E. Walter and L. Pronzato.
Identification of Parametric Models from Experimental Data.
Springer-Verlag, London, 1997.