Non-asymptotic confidence sets for the parameters of dynamical systems: The LSCR approach

Erik Weyer
(ewey@unimelb.edu.au)

Department of Electrical and Electronic Engineering
University of Melbourne, Australia

joint work with

Marco Campi
(marco.campi@ing.unibs.it)

Department of Electrical Engineering and Automation
University of Brescia, Italy
Outline

• Introduction: Dynamical systems, system identification, parameter uncertainty evaluation and confidence sets
• A preview example
• The Leave-out Sign-dominant Correlation Regions (LSCR) algorithm
• Simulation example
• Extensions
• Conclusion
System identification: Building models of dynamical systems based on observed data. Common approach: Parameterise the models and estimate the parameters from data.

A model is never a correct description of the real system. When a model is used for e.g. control design or prediction, it is important to know the uncertainties associated with the model, i.e.

How good is the model?
Parameter uncertainty evaluation

Desirable properties of methods for uncertainty evaluation:

**Applicable under general conditions.**
Example: Restrictive assumptions on the noise (e.g. Gaussian) means that the theory is not applicable in many real life situations.

**Non-conservative evaluation of the model uncertainties.**
Example: A robust controller looses in performance as the uncertainty in the model used for design increases.
Parameter uncertainty evaluation.

Asymptotic theory

- Gives often sensible results, but not always.
- No guarantee of applicability if the number of data points is small.

Finite sample issues

- There will only be a finite amount of data available for construction of confidence sets and evaluation of the model uncertainty.
- Uncertainty evaluation is more important when the uncertainty is significant, e.g. when the information conveyed by the data is limited.

LSCR produces a guaranteed confidence set for the true system parameters based on any finite data sets under very weak assumptions on the noise.
LSCR algorithm

LSCR can be viewed as an estimation method which delivers a set of models rather than a single model.
Outline

• Introduction: Dynamical systems, system identification, model uncertainty evaluation and confidence sets

• A preview example

• The LSCR algorithm

• Simulation example

• Extensions

• Conclusion
A preview example

System

\[ y_t = b^0 u_t + v_t \]

**Aim:** Generate 15 input data \( u_t, \ t = 1, \ldots, 15 \), and construct a confidence interval \( \hat{\Theta} \) such that

\[ Pr\{b^0 \in \hat{\Theta}\} = 0.8 \]

No information about the noise \( v_t \), other than \( v_t \) being independent of \( u_t \). The noise can be biased.
Input signal

\[ u_t = \begin{cases} 
1 & \text{with probability } 1/2 \\
-1 & \text{with probability } 1/2 
\end{cases} \]

Input and output
Correlation between input and prediction error

Model

\[ \bar{y}_t = bu_t + v_t \]

Predictor and prediction error

\[ \hat{y}_t(b) = bu_t, \quad \epsilon_t(b) = y_t - \hat{y}_t(b) = y_t - bu_t = b^0 u_t + v_t - bu_t, \quad t = 1, \ldots, 15 \]

Correlation

\[ E[u_t \epsilon_t(b)] = (b^0 - b) E u_t^2 + E u_t v_t = b^0 - b \]

Hence

\[ b = b^0 \iff E[u_t \epsilon_t(b)] = 0 \]

Idea: Compute estimates of the correlation using different random subsets of the data. Exclude regions in the parameter space where the estimates are positive or negative too many times.

Note: The noise can be arbitrary since \( E u_t v_t = 0 \) by input design.
Random subsample estimates

Let

\[ f_t(b) = u_t \epsilon_t(b), \quad t = 1, \ldots, 15 \]

Compute 19 (scaled) estimates

\[ g_i(b) = \sum_{t=1}^{15} h_{i,t} f_t(b), \quad i = 1, \ldots, 19 \]

\[ h_{i,t} = \begin{cases} 
1 & \text{with probability } 1/2 \\
0 & \text{with probability } 1/2 
\end{cases} \]

\( h_{i,t} \) determines if \( f_t(b) \) is used when we compute the \( i \)th estimate.
Examples of $g_i(b)$ functions

\[
g_1(b) = f_1(b) + f_3(b) + f_4(b) + f_7(b) + f_8(b) + f_9(b) + f_{12}(b) + f_{15}(b)
\]
\[
g_2(b) = f_2(b) + f_3(b) + f_4(b) + f_8(b) + f_{10}(b) + f_{13}(b) + f_{14}(b)
\]
\[
g_3(b) = f_1(b) + f_5(b) + f_8(b) + f_{10}(b) + f_{11}(b) + f_{14}(b)
\]
\[
g_4(b) = f_1(b) + f_2(b) + f_4(b) + f_6(b) + f_8(b) + f_9(b) + f_{10}(b) + f_{12}(b) + f_{15}(b)
\]
\[\vdots\]
The $g_i(b)$ functions

Unlikely that nearly all functions have the same sign for $b = b^0$. Discard regions where at most one function is positive or negative.

The interval $[0.87, 1.12]$ contains $b^0$ with exact probability 0.8.

($b^0 = 1$, $v_t$ i.i.d. normally distributed with mean 0.5 and variance 0.1.)
Experiment repeated 10 times

80% Confidence sets.
A new situation

Data

Less noisy data. \( (b^0 = 1, \text{ i.i.d. normally distributed with mean 0 and variance 0.001.}) \)}
The correlation functions

A smaller confidence interval is obtained. This is achieved automatically without any a-priori knowledge of the noise.
Outline

• Introduction: Dynamical systems, system identification, model uncertainty evaluation and confidence sets
• A preview example
• The LSCR algorithm
• Simulation example
• Extensions of the LSCR algorithm
• Conclusion
The LSCR (Leave-out Sign-dominant Correlation Regions) method

Data generating system

\[ y_t = \frac{B^0(z^{-1})}{A^0(z^{-1})} u_t + v_t \]

\[ A^0(z^{-1}) = 1 + a_1^0 z^{-1} + a_2^0 z^{-2} + \cdots + a_{n_a}^0 z^{-n_a} \]

\[ B^0(z^{-1}) = b_1^0 z^{-1} + b_2^0 z^{-2} + \cdots + b_{n_b}^0 z^{-n_b} \]

Assumptions

- The use can choose \( u_t \), and the choice does not affect \( v_t \).
- The model orders \( n_a \) and \( n_b \) are known. \( A^0(z^{-1}) \) and \( B^0(z^{-1}) \) are co-prime.
Predictors

\[ \hat{y}_t(\theta) = (1 - A(z^{-1}, \theta))y_t + B(z^{-1})u_t \]

\[ A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{n_a} z^{-n_a} \]

\[ B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n_b} z^{-n_b} \]

In linear regression form

\[ \hat{y}_t(\theta) = -a_1 y_{t-1} - \cdots - a_{n_a} y_{t-n_a} + b_1 u_{t-1} + \cdots + b_{n_b} u_{t-n_b} \]

\[ \hat{y}_t(\theta) = \phi_t^T \theta \]

\[ \phi_t = [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-1}, \ldots, u_{t-n_b}]^T \]

\[ \theta = [a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}]^T \]
Example

True system

\[ y_t + a^0 y_{t-1} = b^0 u_{t-1} + w_t + c^0 w_{t-1} \]

can be written as

\[ y_t = \frac{B^0(z^{-1})}{A^0(z^{-1})} u_t + v_t \]

With

\[ A^0(z^{-1}) = 1 + a^0 z^{-1} \]
\[ B^0(z^{-1}) = b^0 z^{-1} \]
\[ v_t = \frac{1 + c^0 z^{-1}}{1 + a^0 z^{-1}} w_t \]
\[ v_t + a^0 v_{t-1} = w_t + c^0 w_{t-1} \]
Predictor

\[
\hat{y}_t(\theta) = (1 - A(z^{-1}, \theta))y_t + B(z^{-1})u_t = -ay_{t-1} + bu_{t-1}
\]

In linear regression form

\[
\hat{y}_t(\theta) = \phi_t^T \theta
\]

\[
\phi_t = \begin{bmatrix} -y_{t-1}, u_{t-1} \end{bmatrix}^T
\]

\[
\theta = \begin{bmatrix} a, b \end{bmatrix}^T
\]
Construction of confidence sets

Input design $u_t$ is i.i.d. and symmetrically distributed around 0.

1. For $t = 1, 2, \ldots, N$ compute the prediction errors

$$\epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = A(z^{-1}, \theta)y_t - B(z^{-1}, \theta)u_t$$

2. For $r = 1, 2, \ldots, n = n_a + n_b$ compute

$$f_{t,r}(\theta) = u_{t-r}\epsilon_t(\theta), \ t = 1, \ldots, N$$

3. Compute $M - 1$ random subsample estimates

$$g_{i,r}(\theta) = \sum_{t=1}^{N} h_{i,t} f_{t,r}(\theta) = \sum_{t=1}^{N} h_{i,t} u_{t-r}\epsilon_t(\theta), \ i = 1, \ldots, M - 1$$

where $h_{i,t} = 0$ or 1 with probability 1/2 each. ($g_{0,r}(\theta) = 0$).

4. Select an integer $q$ and find the region $\hat{\Theta}^r_N$ such that at least $q$ of the $g_{i,r}$ functions are larger than 0 and at least $q$ are smaller than 0.
Probability of the constructed sets

Assume that $Pr\{g_{i,r}(\theta^0) = 0\} = 0$, $i \neq 0$. Then

$$Pr\{\theta^0 \in \hat{\Theta}^r_N\} = 1 - 2q/M$$

Confidence set

The sets $\hat{\Theta}^r_N$ are usually unbounded. A practically useful set is

$$\hat{\Theta}_N = \cap_{r=1}^n \hat{\Theta}^r_N$$

Probability that the confidence set contains the true parameter

$$Pr\{\theta^0 \in \hat{\Theta}_N\} \geq 1 - 2qn/M$$
The shape of the confidence sets

Assume that

1. $A^0(z^{-1})$ is asymptotically stable.
2. $|u_t| \leq U$ for some $U$ and $|v_t| \leq K t^\alpha$ for some $K$ and $\alpha < 1/2$.

Then, for all $\epsilon > 0$

$$Pr\{\exists N(\epsilon) | \hat{\Theta}_N \subseteq \{\theta : ||\theta - \theta^0|| \leq \epsilon\} \; \forall N > N(\epsilon) \} = 1$$

That is, there exists a realisation dependent $N(\epsilon)$ such that the confidence set is included in an $\epsilon$ neighbourhood of $\theta^0$ for all $N > N(\epsilon)$. 
Connection with Instrumental Variable (IV) methods

Main idea behind IV methods: The prediction error should be uncorrelated with past data.

Let $\xi_t = [u_{t-1}, \ldots, u_{t-n}]^T$ be the vector of instrumental variables.

The IV estimate is given by

$$\hat{\theta}_N = \left\{ \theta \text{ such that } \sum_{t=1}^{N} \xi_t \epsilon_t(\theta) = 0 \right\}.$$

In LSCR the confidence set $\hat{\Theta}_N$ is constructed by excluding the regions in parameter space where the components of $\sum_{t=1}^{N} h_{i,t} \cdot \xi_t \epsilon_t(\theta)$ takes on positive or negative values too many times.
Outline

- Introduction: Dynamical systems, system identification, model uncertainty evaluation and confidence sets
- A preview example
- The LSCR algorithm
- Simulation example
- Extensions
- Conclusion
Simulation example

System

\[ y_t = -a^0 y_{t-1} + b^0 u_{t-1} + \tilde{v}_t, \]

where \( a^0 = -0.7 \) and \( b^0 = 0.3 \).

\( \tilde{v}_t \) is biased lowpass filtered white noise.

\[ \tilde{v}_t = \bar{v}_t + 0.5 \]

\[ \bar{v}_t = c_0 \bar{v}_{t-1} + \sqrt{1 - c_0^2} w_t \]

where \( c_0 = 0.5 \) and \( w_t \) is i.i.d. normally distributed with 0 mean and variance 0.12.

\( u_t \) is i.i.d. and uniformly distributed on \([-\sqrt{3}, \sqrt{3}]\).

We have \( N = 1000 \) data points. Want to find a 95% confidence set for \( a^0 \) and \( b^0 \).
Model

\[ y_t + ay_{t-1} = bu_{t-1} \]

Compute the prediction errors

\[ \epsilon_t(a, b) = y_t + ay_{t-1} - bu_{t-1}, \quad t = 1, \ldots, 1000, \]

and compute \( M = 960 \) subsample estimates

\[
g^1_i(a, b) = \sum_{t=1}^{1000} h_{i,t} \cdot u_{t-1} \epsilon_t(a, b), \quad i = 0, \ldots, 959,
\]
\[
g^2_i(a, b) = \sum_{t=1}^{1000} h_{i,t} \cdot u_{t-2} \epsilon_t(a, b), \quad i = 0, \ldots, 959.
\]

Discard regions where 0 is among the \( q = 12 \) largest and smallest values of \( g^1_i(a, b) \) and \( g^2_i(a, b), j = 1, \ldots M \). Then

\[
Pr\{(a^0, b^0) \in \hat{\Theta}\} \geq 1 - 2 \cdot n \cdot q/M = 1 - 2 \cdot 2 \cdot 12/960 = 0.95
\]
Confidence set obtained by excluding regions where \( g_i^1 \) is positive or negative too many times

\[ \begin{align*}
0 \text{ is among the 12 smallest values of } g_j^1(a,b). \\
+ \text{ 0 is among the 12 largest values of } g_j^1(a,b)
\end{align*} \]
Confidence set obtained by excluding regions where \( g_i^2 \) is positive or negative too many times

- 0 is among the 12 smallest values of \( g_j^2(a, b) \).
- 0 is among the 12 largest values of \( g_j^2(a, b) \)
Blank region: Confidence set using LSCR. \( \diamond - (a^0, b^0) \)

\( x/+ \) 0 is among the 12 smallest/largest values of \( g^1_j(a, b) \).

\( o/\square \) 0 is among the 12 smallest/largest values of \( g^2_j(a, b) \).
Blank region: Confidence set using LSCR. ◊ - \((a^0, b^0)\).

Confidence set concentrates around true parameters as \(N\) increases.

Can also show that asymptotically the confidence region is rectangular and the length of the sides decreases as \(1/\sqrt{N}\).
Outline

• Introduction: Dynamical systems, system identification, model uncertainty evaluation and confidence sets
• A preview example
• The LSCR algorithm
• Simulation example
• Extensions
• Conclusion
Input is given

\[ y_t = G^0(z^{-1})u_t + H^0(z^{-1})w_t \]

\( w_t \) independent, symmetrically distributed around zero. \( u_t \) given.

Model class

\[ y_t = G(z^{-1}, \theta)u_t + H(z^{-1}, \theta)w_t \]

True system is included in the model class

\[ G^0(z^{-1}) = G(\theta^0, z^{-1}), \quad H^0(z^{-1}) = H(\theta^0, z^{-1}) \]

Prediction error

\[ \epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = H^{-1}(z^{-1}, \theta)y_t - H^{-1}(z^{-1}, \theta)G(z^{-1}, \theta)u_t \]

Confidence set for \( \theta^0 \) can be obtained using correlations of the type

\[ f_r^\epsilon(\theta) = \epsilon_{t-r}(\theta)\epsilon_t(\theta), \quad f_s^u(\theta) = u_{t-s}\epsilon_t(\theta) \]
Unmodelled dynamics: Example

True system

\[ y_t = G^0(z^{-1})u_t + v_t \]

\( v \) is independent of \( u \), but otherwise arbitrary.

FIR-System

\[ G^0(z^{-1}) = b_0^0z^{-1} + b_0^2z^{-2} \]

Model: \( G(z^{-1}) = b_1z^{-1} \).

Input design: \( u_t \) iid and symmetrically distributed around 0.

Prediction error: \( \epsilon_t(\theta) = y_t - bu_{t-1} \).

Confidence interval for \( b_1^0 \) can be obtained using

\[ f_t(\theta) = \text{sign} \left( u_{t-1}\epsilon_t(\theta) \right) \]
Extensions to non-linear time series

Two modifications:

1) System inversion instead of prediction.

Example

\[ y_t = y_{t-1}(\theta^0 + w_t), \quad \hat{y}_t(\theta) = \theta y_{t-1} \]

\[ \epsilon_t(\theta_0) = y_{t-1}w_t \text{ is not an independent sequence.} \]

System inversion:

\[ \eta_t(\theta) = y_t/y_{t-1} - \theta \]

2) Second order statistics is in general not sufficient for identification of the true parameters. Higher order statistics is used, e.g.

\[ f_t(\theta) = \eta_{t-1}(\theta)\eta_t^2(\theta) \]
Conclusions: LSCR

- The LSCR method generates confidence sets based on a finite number of data points.

- LSCR constructs confidence sets using random subsamples. It is based on the principle that for the true value of the parameters it is very unlikely that nearly all subsample estimates of particular correlations functions are either negative or positive.

- The LSCR has the following properties
  - The confidence set contains the true parameters with guaranteed non-conservative probability.
  - The probability is guaranteed for any finite size of the data set.
  - The prior information on the noise is reduced to a minimum.
References


