MTA SZTAKI
Hungarian Academy of Sciences
Institute for Computer Science and Control

Undermodelling Detection with Sign-Perturbed Sums

Algo Carè [1,2] Marco Campi [3] Balázs Csáji [2] Erik Weyer [4]

[1]Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands

[2]Institute for Computer Science and Control (SZTAKI), Hungarian Academy of Sciences (MTA), Hungary

[3]Department of Information Engineering (DII), University of Brescia, Italy

[4]Department of Electrical and Electronic Engineering, University of Melbourne, Australia

IFAC World Congress, Toulouse, France, July 10, 2017

Table of contents

Motivations

- SPS (Sign-Perturbed Sums) builds confidence regions around the LS (least squares) estimate of linear regression problems.
- Only mild statistical assumptions are needed, e.g., symmetry.
- Not needed: stationarity, moments, particular distributions.
- SPS has many nice properties (as we will see later), most importantly its confindence regions are exact.
- Regarding the models, the assumption of SPS is that the true system generating the observations is in the model class.
- However, if the model class is wrong, SPS cannot detect it.
- Here, we suggest an extension of SPS, UD-SPS, that still builds exact confidence sets, if the model is correct, but can also detect, in the long run, if the system is undermodelled.

Linear Regression

Consider a standard linear regression problem:

Linear Regression

$$y_t \triangleq \varphi_t^{\mathrm{T}} \theta^* + w_t$$

where

- y_t — output (for time $t = 1, \dots, n$)
- φ_t — regressor (exogenous, d dimensional)
- w_t — noise (independet, symmetric)
- θ^* — true parameter (deterministic, d dimensional)
- $\Phi_n = [\varphi_1, \dots, \varphi_n]^{\mathrm{T}}$ — skinny and full rank

Least Squares

Given: a sample, $\mathcal{Z}$, of size n of outputs $\{y_t\}$ and regressors $\{\varphi_t\}$

A classical approach is to minimize the least squares criterion

$$\mathcal{V}(\theta \mid \mathcal{Z}) \triangleq \frac{1}{2}\sum_{t=1}^{n}(y_t - \varphi_t^{\mathrm{T}}\theta)^2.$$

The least squares estimate (LSE) can be found by solving

Normal Equation

$$\nabla_\theta \mathcal{V}(\hat{\theta}_n \mid \mathcal{Z}) = \sum_{t=1}^{n}\varphi_t(y_t - \varphi_t^{\mathrm{T}}\hat{\theta}_n) = 0$$

Confidence Ellipsoids

LSE is asymptotically normal (under some technical conditions)

$$\sqrt{n}\,(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2 R^{-1}) \quad \text{as} \quad n \to \infty,$$

where R is the limit of $R_n = \frac{1}{n}\sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}}$ as $n \to \infty$ (if exists).

Confidence Ellipsoid

$$\widetilde{\Theta}_{n,\mu} \triangleq \left\{ \theta \in \mathbb{R}^d \,:\, (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n (\theta - \hat{\theta}_n) \,\leq\, \frac{\mu\,\hat{\sigma}_n^2}{n} \right\}$$

where $\mathbb{P}(\theta^* \in \widetilde{\Theta}_{n,\mu}) \approx F_{\chi^2(d)}(\mu)$, where $F_{\chi^2(d)}$ is the CDF of $\chi^2(d)$, $\hat{\sigma}_n^2 \triangleq \frac{1}{n-d} \sum_{t=1}^{n} (y_t - \varphi_t^{\mathrm{T}} \hat{\theta}_n)^2$, is an estimate of σ^2.

Reference and Sign-Perturbed Sums

Let us introduce a reference sum and $m - 1$ sign-perturbed sums.

Reference Sum

$$S_0(\theta) \triangleq R_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t (y_t - \varphi_t^{\mathrm{T}} \theta)$$

Sign-Perturbed Sums

$$S_i(\theta) \triangleq R_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} (y_t - \varphi_t^{\mathrm{T}} \theta)$$

for $i = 1, \dots, m - 1$, where $\alpha_{i,t}$ $(t = 1, \dots, n)$ are i.i.d. random signs, that is $\alpha_{i,t} = \pm 1$ with probability $1/2$ each (Rademacher).

Intuitive Idea: Distributional Invariance

Recall: $\{w_t\}$ are independent and each w_t is symmetric about zero.

Observe that, if $\theta = \theta^*$, we have $(i = 1, \dots, m-1)$

Distributional Invariance

$$S_0(\theta^*) = R_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t w_t$$

$$S_i(\theta^*) = R_n^{-\frac{1}{2}} \sum_{t=1}^{n} \varphi_t \, \alpha_{i,t} w_t$$

Consider the ordering $\|S_{(0)}(\theta^*)\|^2 \prec \cdots \prec \|S_{(m-1)}(\theta^*)\|^2$

Note: relation "$\prec$" is the canonical "$<$" with random tie-breaking

All orderings are equally probable! (they are conditionally i.i.d.)

Intuitive Idea: Reference Dominance

What if $\theta \neq \theta^*$?

In fact, the reference paraboloid $\|S_0(\theta)\|^2$ increases faster than $\{\|S_i(\theta)\|^2\}$, thus will eventually dominate the ordering.

Intuitively, for "large enough" $\|\tilde{\theta}\|$, where $\tilde{\theta} \triangleq \theta^* - \theta$

Eventual Dominance of the Reference Paraboloid

$$\left\| \sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}} \tilde{\theta} + \sum_{t=1}^{n} \varphi_t w_t \right\|_{R_n^{-1}}^2 > \left\| \sum_{t=1}^{n} \pm \varphi_t \varphi_t^{\mathrm{T}} \tilde{\theta} + \sum_{t=1}^{n} \pm \varphi_t w_t \right\|_{R_n^{-1}}^2$$

with "high probability" (for simplicity $\pm$ is used instead of $\{\alpha_{i,t}\}$).

Non-Asymptotic Confidence Regions

The rank of $\|S_0(\theta)\|^2$ in the ordering of $\{\|S_i(\theta)\|^2\}$ w.r.t. $\prec$ is

$$\mathcal{R}(\theta) = 1 + \sum_{i=1}^{m-1} \mathbb{I}(\|S_i(\theta)\|^2 \prec \|S_0(\theta)\|^2),$$

where $\mathbb{I}(\cdot)$ is an indicator function.

Sign-Perturbed Sums (SPS) Confidence Regions

$$\widehat{\Theta}_n \triangleq \left\{ \theta \in \mathbb{R}^d \, : \, \mathcal{R}(\, \theta \,) \leq m - q \right\}$$

where $m > q > 0$ are user-chosen integers (design parameters).

Exact Confidence

(A1) $\{w_t\}$ is a sequence of independent random variables.
Each w_t has a symmetric probability distribution about zero.

(A2) The outer product of regressors is invertible, $\det(R_n) \neq 0$.

Exact Confidence of SPS

$$\mathbb{P}\left(\theta^* \in \widehat{\Theta}_n\right) = 1 - \frac{q}{m}$$

for finite samples. Parameters m and q are under our control.

Note that $\|S_0(\hat{\theta}_n)\|^2 = 0$, thus $\hat{\theta}_n \in \widehat{\Theta}_n$, assuming it is non-empty.

Star Convexity

Set $\mathcal{X} \subseteq \mathbb{R}^d$ is star convex if there is a star center $c \in \mathbb{R}^d$ with

$$\forall x \in \mathcal{X}, \forall \beta \in [0,1] : \beta x + (1-\beta) c \in \mathcal{X}.$$

Star Convexity of SPS

$\widehat{\Theta}_n$ is star convex with the LSE, $\hat{\theta}_n$, as a star center

Hint $\widehat{\Theta}_n$ is the union and intersection of ellipsoids containing LSE.

Strong Consistency

(A1) independence, symmetricity: $\{w_t\}$ are independent, symmetric

(A2) invertibility: $R_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}}$ is invertible

(A3) regressor growth rate: $\sum_{t=1}^{\infty} \|\varphi_t\|^4 / t^2 < \infty$

(A4) noise moment growth rate: $\sum_{t=1}^{\infty} \left(\mathbb{E}[w_t^2]\right)^2 / t^2 < \infty$

(A5) Cesàro summability: $\lim_{n\to\infty} R_n = R$, which is positive definite

Strong Consistency of SPS

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ \widehat{\Theta}_n \subseteq B_{\varepsilon}(\theta^*) \right\} \right) = 1,$$

where $B_{\varepsilon}(\theta^*) \triangleq \{\, \theta \in \mathbb{R}^d : \|\theta - \theta^*\| \leq \varepsilon \,\}$ is a norm ball.

Ellipsoidal Outer Approximation

The reference paraboloid can be rewritten as

$$\|S_0(\theta)\|^2 = (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n (\theta - \hat{\theta}_n).$$

From which an alternative description of the confidence region is

$$\widehat{\Theta}_n \subseteq \left\{ \theta \in \mathbb{R}^d \,:\, (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n (\theta - \hat{\theta}_n) \leq r(\theta) \right\},$$

where $r(\theta)$ is the qth largest value of $\{\|S_i(\theta)\|^2\}_{i\neq 0}$.

Ellipsoidal Outer Approximation

$$\widehat{\Theta}_n \subseteq \left\{ \theta \in \mathbb{R}^d \,:\, (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n (\theta - \hat{\theta}_n) \leq r^* \right\}$$

Where r^* can be efficiently computed by a semi-definite program.

Undermodelling

Assume we are given a (finite) sample of input and output data, $\{u_t\}$, $\{y_t\}$, which we model with an FIR system

$$\widehat{y}_t(\theta) \triangleq \varphi_t^{\mathrm{T}}\theta + w_t,$$

where $\varphi_t \triangleq [\, u_{t-1}, \ldots, u_{t-d} \,]^\top$

The true data generation system

$$y_t = \varphi_t^\top \theta^* + e_t + n_t,$$

where e_t is an extra component that can depend on all past inputs $u_{t-d-1}, u_{t-d-2}, \ldots$ and on all past noises $n_{t-1}, n_{t-2} \ldots$.

If $\{e_t\}$ are nonzero, then the SPS confidence regions will still (almost surely) shrink, but around a wrong parameter value.

SPS with Undermodelling Detection

UD-SPS is obtained from SPS by replacing $\{S_i(\theta)\}$ with

$$Q_0(\theta) \triangleq \begin{bmatrix} R_n & B_n \\ B_n^\top & D_n \end{bmatrix}^{-\frac{1}{2}} \frac{1}{n}\sum_{t=1}^{n} \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} (y_t - \varphi_t^\top \theta),$$

$$Q_i(\theta) \triangleq \begin{bmatrix} R_n & B_n \\ B_n^\top & D_n \end{bmatrix}^{-\frac{1}{2}} \frac{1}{n}\sum_{t=1}^{n} \alpha_{i,t} \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} (y_t - \varphi_t^\top \theta),$$

where ψ_t is a vector that includes s extra input values preceding the $\hat{n}_b$ that are included in φ_t, $\psi_t \triangleq [\, u_{t-d-1}, \ldots, u_{t-d-s} \,]^\top$, and

$$B_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \varphi_t \psi_t^\top, \qquad D_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \psi_t \psi_t^\top.$$

The Connection of UD-SPS and SPS

The connection of UD-SPS and SPS can be stated as

Reducing UD-SPS to SPS

The UD-SPS region, $\widehat{\Theta}_n^{\circ}$, for estimating $\theta^* \in \mathbb{R}^d$ can be interpreted as the restriction to a d-dimensional space of a standard SPS region, $\widehat{\Theta}_n'$, that lives in the domain $\{\theta' \in \mathbb{R}^{d+s}\}$.

$\mathbb{R}^{d+s}$ is the d-dimensional identification space augmented with s extra components: $\widehat{\Theta}_n^{\circ}$ can be identified with the first d components of the set $\widehat{\Theta}_n' \cap (\mathbb{R}^d \times \{0\}^s)$.

UD-SPS with Correct System Specifications

Theorem (Exact Confidence of UD-SPS)

If the FIR system is correctly specified, then

$$\mathbb{P}\{\, \theta^* \in \widehat{\Theta}_n^o \,\} \;=\; 1 - q/m.$$

Theorem (Strong Consistency of UD-SPS)

If the FIR system is correctly specified, then (under some technical conditions) for all $\varepsilon > 0$, we have that

$$\mathbb{P}\left[\bigcup_{\bar{n}=1}^{\infty} \bigcap_{n=\bar{n}}^{\infty} \left\{\widehat{\Theta}_n^o \subseteq B_\varepsilon(\theta^*)\right\}\right] = 1,$$

where $B_\varepsilon(\theta^*)$ denotes an ε-ball centred around θ^*.

UD-SPS in the Presense of Undermodelling

Theorem (Undermodelling Detection)

Assume that the system is undermodelled, that is $\{e_t\}$ are nonzero (and some technical conditions hold). With the notations

$$\bar{R}' \triangleq \lim_{n\to\infty} \begin{bmatrix} R_n & B_n \\ B_n^\top & D_n \end{bmatrix}, \quad \bar{E}' \triangleq \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} \mathbb{E}[e_t],$$

if the following detectability condition holds

$$\bar{R}'^{-1} \bar{E}' \notin \mathbb{R}^{\hat{n}_b} \times \{0\}^s,$$

then

$$\mathbb{P}\left[\bigcup_{\bar{n}=1}^{\infty} \bigcap_{n=\bar{n}}^{\infty} \left\{ \widehat{\Theta}_n^o = \emptyset \right\} \right] = 1.$$

Numerical Experiments

Consider the following ARX(1,1) data generating system

$$y_t \;=\; a^* y_{t-1} + b^* u_{t-1} + n_t,$$

with zero initial conditions, where $a^* = 0.5$ or 0.15 or 0 (see later), $b^* = 1$, $\{n_t\}$ are i.i.d. Laplacian with mean 0 and variance 0.1.

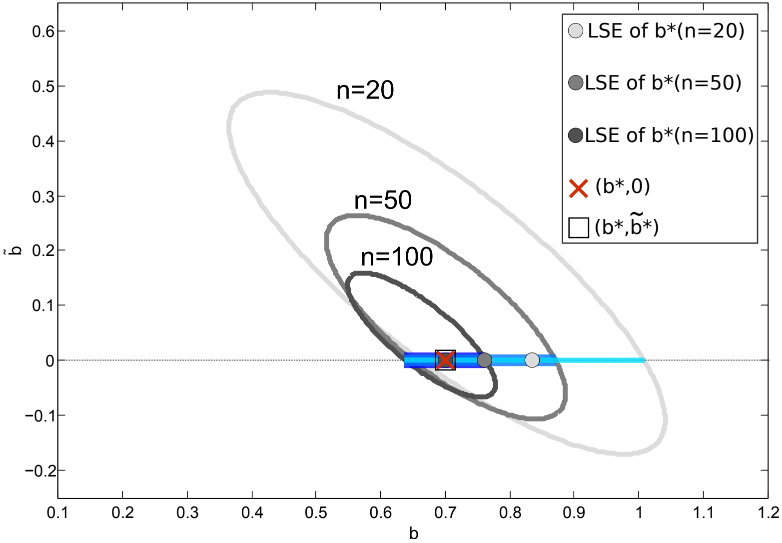
The input signal is generated as $u_t = 0.75u_{t-1} + v_t$, where $\{v_t\}$ are i.i.d. standard normal random variables.

The user-chosen predictor is an FIR(1) model

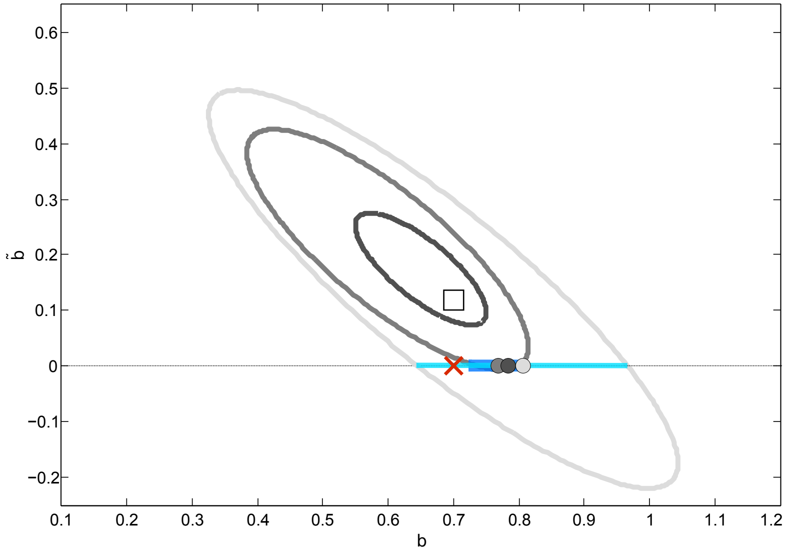
$$\widehat{y}_t(\theta) \;= \varphi_t^\top \theta = b\, u_{t-1},$$

that is, the autoregressive part is missing, $\theta = [\,b\,]$ is the model parameter, and $\varphi_t = [\,u_{t-1}\,]$ is the regressor at time t.

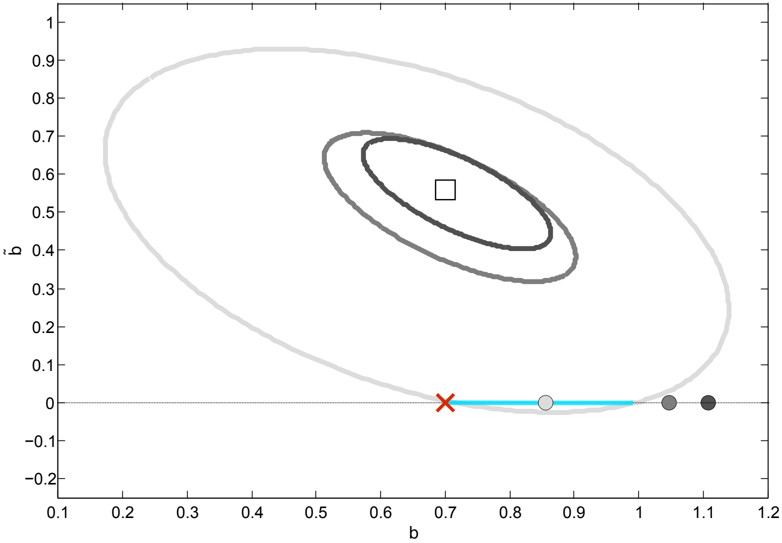
95% UD-SPS Confidence Intervals, $a^* = 0$



95% UD-SPS Confidence Intervals, $a^* = 0.15$



95% UD-SPS Confidence Intervals, $a^* = 0.5$



Summary and Conclusions

- SPS (Sign-Perturbed Sums) is a powerful finite sample system identification method that builds exact, star convex, strongly consistent confidence regions for linear regression problems.
- SPS also has efficient ellipsoidal outer-approximations.
- However, SPS cannot detect if the model class is wrong.
- Here, we suggested an extension of SPS, called UD-SPS, that still guarantees exact and strongly consistent confidence regions if the model order is correctly specified.
- Furthermore, it can detect, in the long run, if the model is undermodelled (detection = empty confidence region).
- There is a strong connection between SPS and UD-SPS.
- The theoretical results were also confirmed by numerical experiments: FIR models of ARX systems were studied.

Thank you for your attention!

✉ balazs.csaji@sztaki.mta.hu