IMPROVEMENT TO THE LEAVE-OUT SIGN-DOMINANT CORRELATION REGION METHOD

Juan Carlos Jarur, and Ricardo A. Rojas
Department of Electronics Engineering, Universidad Técnica Federico Santa Maria
Valparaiso
Chile (corresponding author e-mail:jjarur@gmail.com)

ABSTRACT
We propose two refinements to the LSCR (Leave-Out Sign-Dominant Correlation Regions) method to improve the construction of confidence regions for parameters of identified models with a guaranteed probability. The LSCR method holds for any finite number of data points without using asymptotic theory, and previous knowledge on the noise affecting the data is reduced to a minimum. We prove the exact probability for the general confidence region instead of just a probability bound as it is the current situation. We also detect an accuracy problem when using a straightforward implementation of the LSCR algorithm and provide a solution by taking advantage of the new found exact probability expression. The theory is validated with empirical results based on Monte Carlo simulations on open loop and closed loop systems.

KEY WORDS
Correlation Functions, Confidence Regions, LSCR, Prediction Error.

1. Introduction

In many applications, uncertainty evaluation and parameter validation are delivered based on the asymptotic theory of system identification. Expressions for uncertainty evaluation, confidence ellipsoids, and asymptotic covariance of estimated quantities are developed to compute confidence intervals to assess the reliability of particular estimates obtained from an observed data set [6]. Most of the expressions are asymptotic in the number of data points and some theory does not contain information about how large N has to be for results to be applicable. The previews work [7] shows that the uncertainty regions constructed through the asymptotic theory can be unreliable giving sensible results in certain situations. Also, some operating conditions are necessary for the asymptotic theory to exhibit a robustness that makes it reliable with a finite number of data [8]. In general, the asymptotic theory is correct only when the number of data tends to infinity. Moreover, in practical applications the number of data points available for an identification procedure is finite and even scarce [6],[10]. In these applications we find that certain restrictive assumptions are difficult to assess, e.g. the assumption of bounded Gaussian noise [9].

On the positive side, asymptotic theory tends to provide confidence regions which are easily constructed from the data, giving regions which are reliable and with a bounded description of uncertainty.

An advance towards working with finite data were the results obtained in [11],[12], which hold true without using asymptotic theory. Another important contribution in this sense is Bootstrap [13]. The development of Bootstrap with computer speed improvements made it possible to calculate uncertainty regions for parameter identification problems [14]. The main idea behind bootstrap is computing Monte Carlo simulations to judge the uncertainty in estimated parameters. A simulation study of the bootstrap confidence intervals for estimating parameters, assuming some properties for the residuals, is introduced in [14],[15].

The LSCR (Leave-Out Sign-Dominant Correlation Regions) method developed in [1]-[4] offers another approach to construct confidence sets for general linear systems based on a finite number of data points. An improvement over the Bootstrap method is that LSCR provides confidence sets to which the true system parameter belongs with an exact guaranteed probability. No probability bounds and no conservativeness are introduced in the found sets. However, in order to obtain a practical useful confidence set (or region), denoted by $\hat{\Theta}$, around the true parameter a probability bound (not an exact one) is introduced. A nice feature of the LSCR method is that much knowledge on the noise level is not necessary when constructing confidence regions. Most of the theory is developed in [2]-[4], providing a method for model quality evaluation given any data set size, even in presence of unmodeled dynamics [4].

The LSCR method needs to satisfy two properties to guarantee its performance: the confidence region must have guaranteed probability, and the region must be bounded and concentrate around the real parameter, denoted by θ^0, as the number of data points increases. We prove, in this contribution that in some cases presented in [1]-[3] the first property does not hold.

Moreover, the probability bound presented in [1]-[4] is very inaccurate, and even negative for low probability confidence regions.

Our first contribution is a new expression giving the exact probability for the event $\theta^0 \in \hat{\Theta}$. We demonstrate that this new result is essential to a good implementation of the LSCR method. Next, we present a generalization of this new result.

Our second contribution has to do with a problem concerning a direct implementation of the LSCR method presented in [1]-[3]. After presenting the problem, which appears while applying LSCR to slow dynamic systems and different signal-to-noise ratios, we offer a solution.

The main ideas of LSCR and our improvements to LSCR implementations are developed for the AR, ARMA and Box & Jenkins models (Closed loop) taken from [1]-[3] and within a Monte Carlo framework.

1.1 Contents of the present paper

In section 2 we present an overview of the LSCR method exposing their most important properties and explaining the construction procedure of $\hat{\Theta}$ given in [1]-[3].

In section 3, we present a new general expression for the exact probability that the real parameter θ^0 belongs to the confidence region $\hat{\Theta}$. The theorem is proved in the Appendix.

Finally in section 4, and using the new result, we show that for certain cases, the probability of θ^0 belonging to $\hat{\Theta}$ does not hold when implementing the LSCR algorithm from a direct interpretation of [1]-[4].

2. The LSCR Method

This section presents the most important features of LSCR method [1]-[3]. The name of the LSCR method comes from the computation of empirical correlation functions and to use those of them for which the parameter combinations, defined by θ, do not produce either too many positive or too many negative values. This principle is based on the fact that the correlation functions evaluated for the real parameter value are sums of zero mean random variables and, therefore, it is unlikely that almost all of them will be positive or almost all of them will be negative [1].

2.1 An Example [1], [2]

In order to introduce the LSCR method we start by describing a simple example as a generating system

$$y_t + \theta^0 y_{t-1} = w_t \qquad (1)$$

where $\theta^0 = 0.2$. We are assuming w_t is an independent process with a symmetrical distribution around zero. The noise w_t can have any distribution and level, since they are not used in the LSCR algorithm. The idea is to form a confidence region for θ^0 using 9 data points generated with (1). See Figure 1.

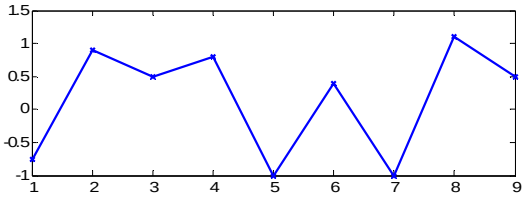

Fig. 1. Available data points for the example.

We now rewrite the system as a model with generic parameter θ

$$y_t + \theta y_{t-1} = w_t \qquad (2)$$

Next, we compute the prediction error for $t = 1,...,8$ associated with the model and all possible values of $|\theta| < 1$ and calculate

$$\varepsilon_t(\theta) = y_t - \theta y_{t-1} \quad \rightarrow \quad f_{t-1}(\theta) = \varepsilon_{t-1}(\theta) e_t(\theta), \quad t = 2,...,8$$

Then we take some of these $f_{t-1}(\theta)$ functions to form empirical correlation functions $E[\varepsilon_{t-1}(\theta) e_t(\theta)]$ and noting that for $\theta = \theta^0$ we have $E[\varepsilon_{t-1}(\theta^0) e_t(\theta^0)] = E[w_{t-1} w_t] = 0$. Therefore, we expect the empirical estimates to be zero mean random variables for $\theta = \theta^0$. Thus, we form 8 empirical estimates of the correlation functions $g_i(\theta)$, as follows:

$$g_i(\theta) = \sum_{k \in I_i} f_k(\theta) \quad i = 1,...,8 \qquad (3)$$

where the I_i sets are subsets of $\{1,...,8\}$ represented by a 1 in the incident matrix (4) (See Appendix A.5 in [2]). The number 1 means that the element is in the set.

$$R(8) = \begin{bmatrix} 1&1&0&1&1&0&0&0 \\ 1&0&1&1&0&1&0&0 \\ 0&1&1&0&1&1&0&0 \\ 1&1&0&0&0&1&1&0 \\ 1&0&1&0&1&0&1&0 \\ 0&1&1&1&0&0&1&0 \\ 0&0&0&1&1&1&1&0 \\ 0&0&0&0&0&0&0&0 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \\ I_8 \end{bmatrix} \qquad (4)$$

The last element I_8, is an exceptional set forming $g_8(\theta) = 0$ [2]. The incident matrix R is generated using subsets of $I = \{1,...,8\}$ which is a group with respect to the symmetric difference, i.e $(I_i \cup I_j) - (I_i \cap I_j) \in R$ (See Appendix A.5 in [2]). The calculated $g_i(\theta)$ functions,

$i = 1,...,8$ are displayed in Figure 2.

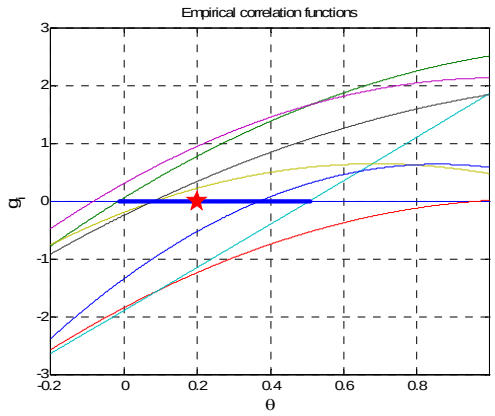

Fig. 2. The resulting interval and the $g_i(\theta)$ functions.

By considering Figure 2, we conclude that the $g_i(\theta)$ functions have a tendency to intersect the θ axis near the real parameter θ^0, and all the empirical correlation functions $g_i(\theta^0)$ take on positive or negative value with equal probability. To construct the confidence interval, we expect that there are at least two $g_i(\theta^0)$ functions greater than zero and at least two $g_i(\theta^0)$ functions smaller than zero at the same θ value. The resulting confidence region for θ^0 is therefore $[-0.01, 0.5]$ with a probability $1-2\cdot 2/8 = 0.5$ (see *Theorem 1* below) to contain the true parameter θ^0.

2.2 The Data generating system

Consider the general linear system presented by

$$y_t = G(q^{-1},\theta^0)u_t + H(q^{-1},\theta^0)w_t$$

We assume that the noise w_t and the input signal u_t are independent stochastic processes, where w_t has zero mean and it has a symmetric distribution around zero. No *a-priori* knowledge of the noise level is assumed. $G(q^{-1},\theta^0)$ and $H(q^{-1},\theta^0)$ are stable rational transfer functions. $H(q^{-1},\theta^0)$ is monic and has a stable inverse [1]. The input signal sequence u_t is a Gaussian stochastic process with zero mean.

We present the system structure taking a full-order model class, as follows

$$y_t = G(q^{-1},\theta)u_t + H(q^{-1},\theta)w_t$$

The parameter space, defined by Θ, is restricted to a set $\theta \in \Theta$ such that $H(q^{-1},\theta)$ is monic and $G(q^{-1},\theta)$, $H(q^{-1},\theta)$, and $H^{-1}(q^{-1},\theta)$ are all stable.

2.3 Construction of confidence regions [1]-[3]

In this section we present the algorithm for general linear systems, describing procedures to determine confidence sets Θ_r^ε and Θ_s^u based on empirical autocorrelations of prediction error $\varepsilon_t(\theta)$ and empirical cross-correlations of $\varepsilon_t(\theta)$ and the input u_t, respectively. Finally, we present a general confidence region $\hat{\Theta}$ by taking the intersection of Θ_r^ε and Θ_s^u sets.

***Procedure for the construction of Θ_r^ε and Θ_s^u:* [1]-[3]**

Step1: Compute the prediction error

$$\varepsilon_t(\theta) = H^{-1}(q^{-1},\theta)y_t - H^{-1}(q^{-1},\theta)G(q^{-1},\theta)u_t$$

for a finite number of values of t, say $t = 1,2,...,K$.

Step2: Select integer $r \geq 1$ or $s \geq 1$. For $t = 1+r,...,N+r = K$ or $t = 1+s,...,N+s = K$, compute

$$f^\varepsilon_{t-r,r}(\theta) = \varepsilon_{t-r}(\theta)\varepsilon_t(\theta) \quad for\ \Theta_r^\varepsilon$$
$$f^u_{t-s,s}(\theta) = u_{t-s}(\theta)\varepsilon_t(\theta) \quad for\ \Theta_s^u$$

Step3: Let $I = \{1,...,N\}$ and consider an incident matrix R constructed of subsets $I_i \subseteq I$, $i = 1,...,M$, forming a group under the symmetric difference operation, i.e $(I_i \cup I_j) - (I_i \cap I_j) \in R$ [5], compute

$$g^\varepsilon_{i,r}(\theta) = \sum_{k\in I_i} f^\varepsilon_{k,r}(\theta), \quad i = 1,...,M \quad for\ \Theta_r^\varepsilon$$
$$g^u_{i,s}(\theta) = \sum_{k\in I_i} f^u_{k,s}(\theta), \quad i = 1,...,M \quad for\ \Theta_s^u$$

Step 4: Select an integer q in the interval $[1,(M+1)/2]$ and find the region Θ_r^ε (Θ_s^u) such that at least q of the $g^\varepsilon_{i,r}(\theta)$ ($g^u_{i,s}(\theta)$) functions are bigger than zero *and* at least q are smaller than zero.

The above procedure for the construction of Θ_r^ε is the same used in the previous example, where we had $H^{-1}(q^{-1},\theta) = 1+\theta q^{-1}$, $G(q^{-1},\theta) = 0$, and $K = 8$, $N = 7$, $r = 1$, $M = 8$, and $q = 2$.

Theorem 1 [2] gives the exact probability for the true parameter to belong to Θ_r^ε or Θ_s^u sets constructed above.

Theorem 1. Assume that the variables w_t and $w_t u_\tau$ admit densities and that w_t is symmetrically distributed around zero. Then the sets Θ_r^ε and Θ_s^u are such that

$$\Pr\{\theta^0 \in \Theta_r^\varepsilon\} = 1 - 2q/M \quad (5)$$
$$\Pr\{\theta^0 \in \Theta_s^u\} = 1 - 2q/M \quad (6)$$

This theorem is shown in Appendix A.1 of [2]. The respective sets are a non-asymptotic confidence region for

θ^0. However, these sets are not useful because Θ_r^ε and Θ_s^u are unbounded in some directions of the parameter space. A practical confidence set $\hat{\Theta}$ can be obtained by intersecting a number of the sets Θ_r^ε and Θ_s^u [2], i.e.

$$\hat{\Theta} = \bigcap_{r=1}^{n_\varepsilon} \Theta_r^\varepsilon \bigcap_{s=1}^{n_u} \Theta_s^u \tag{7}$$

The election of n_ε and n_u depends on the model class under consideration. In order to obtain well shaped confidence sets, the correct choice of n_ε and n_u is discussed in [1][2]. Theorem 2 [2] is immediate from theorem 1, and it provides a *probability bound* for the true parameter to belong to $\hat{\Theta}$.

Theorem 2. Under the assumptions of Theorem 1,

$$\Pr\{\theta^0 \in \hat{\Theta}\} \geq 1 - (n_\varepsilon + n_u) 2q / M \tag{8}$$

According to [2], the inequality is due to the events $\{\theta^0 \notin \Theta_r^\varepsilon\}$, $\{\theta^0 \notin \Theta_s^u\}$, $r = 1,...,n_\varepsilon$, $s = 1,...,n_u$ may be overlapping.

3. Exact Probability

We present our main result. It gives the exact probability for the real parameter θ^0 belonging to the confidence region $\hat{\Theta}$ instead of the lower bound (8). We prove a general expression, which takes into account different q and M values for each subsets Θ_r^ε and Θ_s^u.

Theorem 3. Under the assumptions of Theorem 1,

$$\Pr\{\theta^0 \in \hat{\Theta}\} = \left[1 - \frac{2q}{M}\right]^{(n_\varepsilon + n_u)} \tag{9}$$

Or using different parameter M and q for each subsets

$$\Pr\{\theta^0 \in \hat{\Theta}\} = \prod_{i=1}^{n_\varepsilon + n_u} \left[1 - \frac{2q_i}{M_i}\right] \tag{10}$$

We prove (9) and (10) in Appendix A.1 of this paper.

A Monte Carlo simulation on the behavior of the real parameter θ^0 belonging to the confidence region $\hat{\Theta}$ for different probability levels is presented in Figure 3. We calculate the relative frequency m/E, where m is the number of times that the event $\theta^0 \in \hat{\Theta}_i$ occurs $i = 1,...,5000$ $(E = 5000)$ for different realizations of w_t and desired probabilities in (8) and (9). The simulations were developed considering a first order ARMA system taken from [1]-[3]:

$$y_t + a^0 y_{t-1} = w_t + c^0 w_{t-1} \tag{11}$$

where $a^0 = -0.5$, $c^0 = 0.2$ and w_t is an independent sequence of zero mean Gaussian random variables with unit variance. We use a model class with generic parameters $y_t + a y_{t-1} = w_t + c w_{t-1}$ subject to $|a| < 1$ and $|c| < 1$ for stability. Next, we compute the prediction errors

$$\varepsilon_t(a,c) = \frac{1 + aq^{-1}}{1 + cq^{-1}} y_t \tag{12}$$

then, we follow the procedure for construction of the Θ_r^ε computing *step 2* and *step 3* with $r = 1,2$.

$$f_{t-1,1}^\varepsilon(a,c) = \varepsilon_{t-1}(a,c)\varepsilon_t(a,c), \quad t = 2,...,1024$$
$$f_{t-2,2}^\varepsilon(a,c) = \varepsilon_{t-2}(a,c)\varepsilon_t(a,c), \quad t = 3,...,1025$$

and

$$g_{i,1}^\varepsilon(a,c) = \sum_{k \in I_i} f_{k,1}^\varepsilon(a,c), \quad i = 1,...,1024$$
$$g_{i,2}^\varepsilon(a,c) = \sum_{k \in I_i} f_{k,2}^\varepsilon(a,c), \quad i = 1,...,1024$$

We run this process for $E = 5000$ realizations and count the number of times m that the event $\theta^0 \in \hat{\Theta}_i$ occurs.

Figure 3 presents the statistics m/E for the different values q that can take, $q \in [1,(M+1)/2]$ (see step 4 in construction of Θ_r^ε and Θ_s^u). We see that equation (9) (solid line) coincides with relative frequencies (stars) for the whole range. On the contrary, equation (8) (dotted line) is a good approximation only for high relative frequencies, and it can even take on negative values. Hence, it is clear that equation (8) is a poor approximation for the exact probability of $\theta^0 \in \hat{\Theta}$.

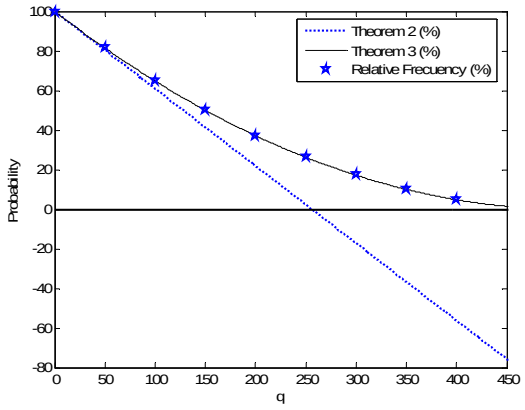

Fig. 3. Comparison of equations (8), (9), and relative frequency m/E.

Next, we show that the equation (8) corresponds to a 2-term Taylor's expansion of equation (9) around $2q/M = 0$. The two first terms of Taylor's expansion evaluated in $q_0 = 0$, give

$$\Pr\{q\} \approx \Pr\{q_0\} + \left.\frac{d\Pr\{q\}}{dq}\right|_{q=q_0} (q-q_0) + \ldots \approx 1 - (n_\varepsilon + n_u)2q/M \quad (13)$$

Equation (13) explains why (8) is a good approximation of (9) only for $2q/M \ll 1$ (See Figure 3).

4. A LSCR Implementation

In this section, we use equation (9) to determine whether the LSCR method is applied correctly, so that it can keep its valuable properties. We do that by checking different probabilities for the event $\theta^0 \in \hat{\Theta}$ and compare with the results given by LSCR *with* and *without* taking into account the prediction error transient.

4.1 ARMA example

In this section we repeat the same Monte Carlo simulation of section III, using $E = 5000$ realizations. We only change the zero in the ARMA system (11), from $c^0 = 0.2$ to $c^0 = 0.9$ (still giving a stable $H^{-1}(q^{-1}, \theta)$). We empirically verify that when the prediction error transient in (12) is not discarded, the probability properties (5), (6) and (9) do not hold. This problem increases as the data points decrease and gets worse when the prediction error transfer function has slow dynamics. We stress that we are not referring to the output y_t transient.

Figure 4(a) shows the relative frequencies given by the Monte Carlo simulations when the residual transient (12) using $N = 1024$ points (stars) are not discarded. In Figure 4 (b) the problem gets worse when the number of data points is lower ($N = 256$), resulting in a smaller relative frequency (stars). The square patrons in Figure 4 (a) and (b) represent the outcomes *without* the residual transient.

Figure 4 shows that the effect of including the prediction error transient can produce large errors when obtaining the confidence region.

In Figures 5, we built two $(1 - 2 \cdot 12/1024)^2 = 0.9532$ confidence regions for $(a^0, c^0) \in \hat{\Theta}$ using $N = 1024$ under two conditions: (a) represents the region *without* residual transient; (b) represents the region *with* the residual transient. Even though the confidence region of Figure 5 may seem well shaped around the true parameter, it may not guarantee the probability needed by the user. The real parameter (a^0, c^0) is represented with a star and the confidence set with a dotted region.

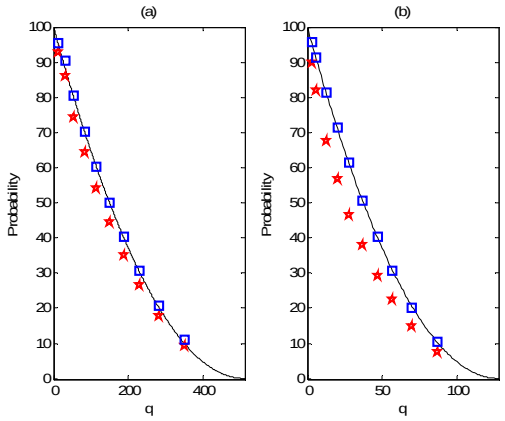

Fig. 4. (a) Simulation for N=1024 data points; (b) Simulation for N=256 data points. □ relative frequency without residual transient, ★ relative frequency with residual transient.

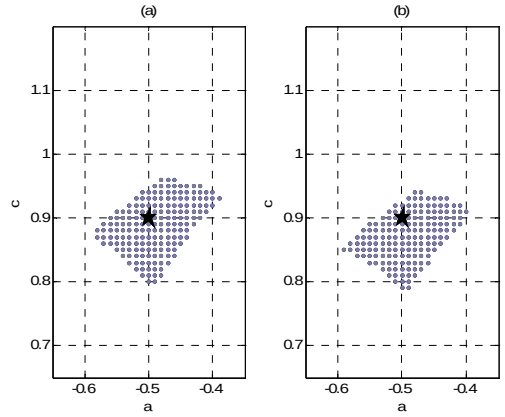

Fig. 5. 95% confidence region: (a) LSCR without transient; (b) LSCR with transient. ★ = real parameters, · parameters of the confidence region.

4.2 A closed Loop example

We now check the implementation of LSCR method using the example presented in [1]-[2] and sketched in Figure 6.

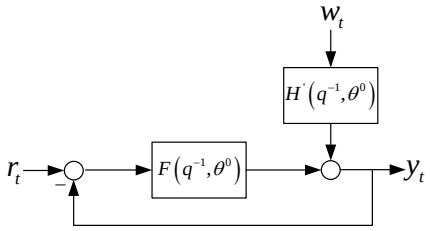

Fig 6. The closed loop system.

$$F(q^{-1},\theta^0) = \frac{b^0 q^{-1}}{1 + a^0 q^{-1}}, \quad a^0 = -0.7 \quad b^0 = 0.3$$
$$H'(q^{-1},\theta^0) = 1 + h^0 q^{-1}, \quad h^0 = 0.5 \quad (14)$$

w_t and the reference signal r_t are independent stochastic processes with variances 1 and 10^{-6}, respectively. We follow the process described in [1][2] taking 2050 data points of r_t and y_t from the system in Figure 6 . We use the prediction error corresponding to Figure 6

$$\varepsilon_t(\theta) = \frac{1+(a+b)q^{-1}}{(1+aq^{-1})(1+hq^{-1})} y_t - \frac{bq^{-1}}{(1+aq^{-1})(1+hq^{-1})} r_t \qquad (15)$$

The final region is constructed using the steps described in Section 2, where we compute

$$g^{\varepsilon}_{i,r}(\theta) = \sum_{k\in I_i} \varepsilon_{k-r}(\theta)\varepsilon_k(\theta), \quad r = 1,2,3. \qquad (16)$$

Thus, we calculate equation (15) and (16) for $E = 5000$ realization r_t and w_t using different seeds. Then, based on the same experimental process explained in the *ARMA example* we count the number of times the event $\theta^0 \in \hat{\Theta}$ occurs. Next we calculate the relative frequency for different probabilities considering that $n_\varepsilon = 3$, $n_u = 0$, $M = 2048$, and $q \in [1,(M+1)/2]$. Figure 7 (a) shows that the relative frequencies (circles) coincide with those given by equation (9). We now change the input signal levels of w_t and r_t while keeping the system dynamics. We do that by simply reversing the variances of w_t and r_t respectively to 10^{-6} and 1, to improve the signal-to-noise ratio presented in the original example. We obtain the Figure 7 (b) where the point mark with stars represent the relative frequencies obtained with (15) *including* transient and the point mark with squares represent the relative frequencies of (15) *without* the transient. We stress that the amount of empirical correlation changes the probability, but the election of auto correlation or cross correlation, restricted by $n_\varepsilon + n_u$, do not change $\Pr\{\theta^0 \in \hat{\Theta}\}$ (See equation (9)).

An important claim of LSCR method is the ability to give a region of guaranteed probability despite that no *a-priori* knowledge on the noise level is assumed [1][2]. However, in some cases a straightforward implementation of LSCR method gives a region without a guaranteed probability. In figure 7(b) the stars show how the prediction error transient affects the results depending on the noise levels.

Now, in order to present another implementation of the LSCR method we set the variances of signals r_t and w_t to 1 and 10^{-3}, respectively, and take 2350 data points from the system in Figure 6. We then calculate the prediction error (15) and discard the first 300 points of $\varepsilon_t(\theta)$. Finally, to construct a well shaped region around the true parameter, compute ($u_t = r_t$)

$$g^{\varepsilon}_{i,r}(\theta) = \sum_{k\in I_i} \varepsilon_{k-r}(\theta)\varepsilon_k(\theta), \quad r = 1.$$
$$g^{u}_{i,s}(\theta) = \sum_{k\in I_i} r_{k-s}(\theta)\varepsilon_k(\theta), \quad s = 1,2.$$

Assuming that the closed loop system is stable for all $\theta \in \Theta$ we exclude the parameters where θ is within the q=35 smallest or largest values for each of the three correlations calculated above. The confidence region in Figure 8 (a) is obtained with 90.09% of confidence, *without* the transient. Figure 8 (b) shows the region $\hat{\Theta}$ obtained when the transient in (14) is *included*. Using a Monte Carlo simulation we obtain that the probability error in Figure 8 (b) is 1.88%.

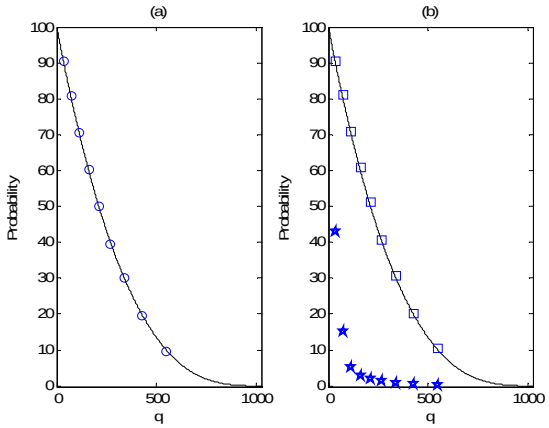

Fig. 7. Monte Carlo simulation (a) w_t and r_t with variance 1 and 10^{-6}; (b) w_t and r_t with variance 10^{-6} and 1. ○ Relative frequency with transient, ★ Relative frequency with transient, □ Relative frequency without transient.

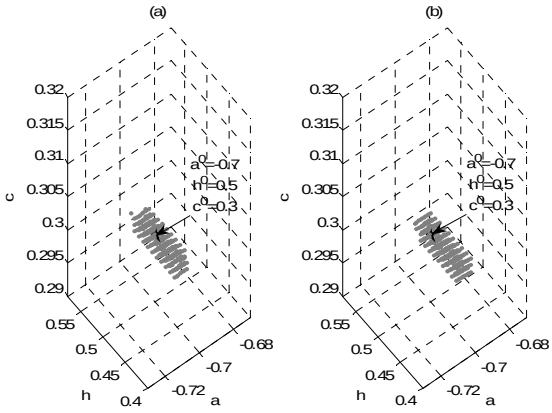

Fig. 8. 90.09% confidence region: (a) LSCR without transient; (b) LSCR with transient. ★ = real parameters, ▪ parameters of the confidence region.

4.3 Construction of the regions

The dynamic system generating the prediction error (17) will in general exhibit a transient in its response even, as

it is the present case, for zero initial conditions. This transient can be very disturbing for LSCR in the presence of slow natural modes, and some signal-to-noise ratios. As a consequence of this transient, the statistics $g^{\varepsilon}_{i,r}(\theta)$ and $g^{u}_{i,s}(\theta)$ may discard some possible parameters θ, thus resulting in a smaller region $\hat{\Theta}$ than it should be. This result is made evident by the small relative frequencies obtained when using the residual transient.

In order to avoid the above problem, we redefined the steps 1 and 2 in the procedure of constructing confidence regions for Θ^{ε}_{r} and Θ^{u}_{s} as follows:

Step1: Compute the prediction error

$$\varepsilon_t(\theta) = H^{-1}(q^{-1},\theta)y_t - H^{-1}(q^{-1},\theta)G(q^{-1},\theta)u_t \quad (17)$$

for a finite number of data, say $t = 1,2,...,K$. K has to be large enough to account for the residual transient removal in (17) and to construct the confidence region.

Step 2: Select integers $r \geq 1$ or $s \geq 1$. For $t = K - M + -\max(n_u, n_\varepsilon) - 2,...,K$, compute

$$f^{\varepsilon}_{t-r,r}(\theta) = \varepsilon_{t-r}(\theta)\varepsilon_t(\theta) \quad for\ \Theta^{\varepsilon}_{r}$$
$$f^{u}_{k-s,s}(\theta) = u_{t-s}(\theta)\varepsilon_t(\theta) \quad for\ \Theta^{u}_{s}$$

then, continue the usual *steps 3* and *4*.

In step 1, it is advisable to obtain the step response of (17) to get an approximation of how many data points have to be discarded in order to develop a sound implementation of the LSCR method.

Remark: An alternative to avoid discarding data points is to properly initialize (17). To this purpose, equation (17) requires *good* initial values of inputs, outputs, and predictions for $t = 0$. To put all initial values to zero is not enough when equation (17) has slow dynamics. The Matlab System Identification Toolbox [16] offers a number of options to deal with the prediction initial state. But it usually implies higher computational cost and the method may not be a viable procedure [2].

5. Conclusion

In this paper, we present two improvements to the LSCR method for constructing non-asymptotic confidence regions over full order linear models, and in the presence of unmodeled dynamics. We have proved a new result capable of giving the exact and general probability for the true parameter θ^0 belonging to the confidence region $\hat{\Theta}$. We also demonstrated that the current probability bound expression corresponds to a 2-term Taylor's expansion of the exact result. This new equation has been found useful to detect a problem with a straightforward implementation of LSCR. We showed that the construction of confidence regions must be obtained after discarding the residual transient of the prediction error or after a proper initialization of the prediction. We have found that this problem gets worse when using small data points, in the presence of slow dynamics, and with some signal-to-noise ratios. All the results have been validated by Monte Carlo simulations.

Appendix

A.1 Proof of (9) (10)

We first follow the proof of equation (5) and (6) in appendix A.1 of [2]:

Note now that $g^{\varepsilon}_{i,r}(\theta^0) = \sum_{k\in I_i} w_k w_{k+r}$ and recall the definition of $g^{\varepsilon}_{i,r}(\theta)$ in the construction process of the set Θ^{ε}_{r}. We define the event $\theta^0 \notin \Theta^{\varepsilon}_{r} = A_r$: Then, $g^{\varepsilon}_{i,r}(\theta) > 0$ for at most $q_r - 1$ selections of I_i *or* it is less than 0 for at most $q_r - 1$ selections.

$$A_r = \left\{\sum_{k\in I_i} w_k w_{k+r} < 0 \text{ for at most } q_r - 1 \text{ selections of } I_i\right\} \cup \left\{\sum_{k\in I_i} w_k w_{k+r} > 0 \text{ for at most } q_r - 1 \text{ selections of } I_i\right\}$$

and the probability of $\Pr\{A_r\}$

$$\Pr\{A_r\} = 2q_r / M_r$$

We may note that there exists a simple relationship between the probability of an event A_r and the probability of the complimentary event $\bar{A}_r = B_r$. We hence define the event $\theta^0 \in \Theta^{\varepsilon}_{r} = B_r$: Then, $g^{\varepsilon}_{i,r}(\theta) > 0$ for at least q_r selections of I_i *and* it is less than 0 for at least q_r selections. Then, the probability for a single subset is

$$1 = \Pr\{\bar{A}_r = B_r\} + \Pr\{A_r\} \quad \Rightarrow \quad \Pr\{B_r\} = 1 - 2q_r / M_r$$

By checking the step 4 in the construction of the set Θ^{ε}_{r}

$$B_r = \left\{\sum_{k\in I_i} w_k w_{k+r} < 0 \text{ at least } q_r \text{ election of } I_i\right\} \cap \left\{\sum_{k\in I_i} w_k w_{k+r} > 0 \text{ at least } q_r \text{ election of } I_i\right\}$$

with B_r, $r = 1,...,n_\varepsilon$.

$$\Pr\{B_1 \sim \theta^o \in \Theta^{\varepsilon}_{1}\} = 1 - 2q_1 / M_1 \ \cdots\ \Pr\{B_{n_\varepsilon} \sim \theta^o \in \Theta^{\varepsilon}_{n_\varepsilon}\} = 1 - 2q_{n_\varepsilon} / M_{n_\varepsilon}$$

where $q_i \neq q_j$, $\forall i \neq j$.

Proposition 1: In general, the occurrence of the event B_1 would affect the probability which we assign to another event B_2. However, it may turn out that, in this case, the conditional probability of B_2 given B_1 is exactly the same as the unconditional probability of B_2, as follows

$$\Pr\{B_2 \mid B_1\} = \Pr\{B_2\}$$

In this case, we say that B_2 and B_1 are independent events. Particularly, the occurrence of B_2 in no way affects the probability of the occurrence of B_1. The event B_2 does not provide information about the event B_1. This important result is known as the multiplication law for independent events

$$\Pr\{B_1 \cap B_2\} = \Pr\{B_1\}\Pr\{B_2\}$$

The only way to affect the probability of $\Pr\{B_1\}$ is by changing the variables q_1 and M_1. So, we are asking for the probability of $\theta^0 \in \hat{\Theta}$, with

$$\hat{\Theta} = \bigcap_{r=1}^{n_\varepsilon} \Theta_r^\varepsilon$$

where we define $B_r \sim \theta^o \in \Theta_r^\varepsilon$, $r = 1,...,n_\varepsilon$.

$$\Pr\{\theta^o \in \hat{\Theta}\} = \Pr\{\theta^o \in \Theta_1^\varepsilon \cap ... \cap \theta^o \in \Theta_{n_\varepsilon}^\varepsilon\}$$

$$\Pr\{\theta^o \in \hat{\Theta}\} = \left(1 - \frac{2q_1}{M_1}\right) \cdots \left(1 - \frac{2q_{n_\varepsilon}}{M_{n_\varepsilon}}\right) \qquad \text{(A.1)}$$

Now we can express (A.1) for a more general case using $\hat{\Theta} = \bigcap_{r=1}^{n_\varepsilon} \Theta_r^\varepsilon \bigcap_{s=1}^{n_u} \Theta_s^u$.

$$\Pr\{\theta^0 \in \hat{\Theta}\} = \prod_{i=1}^{n_\varepsilon + n_u} \left[1 - \frac{2q_i}{M_i}\right] \qquad \text{(A.2)}$$

which simplifies to the basic result when $q_1 = q_2 = ...$ $... = q_{n_\varepsilon + n_u}$ and $M_1 = M_2 = ... = M_{n_\varepsilon + n_u}$.

$$\Pr\{\theta^o \in \hat{\Theta}\} = \left[1 - \frac{2q}{M}\right]^{(n_\varepsilon + n_u)} \qquad \text{(A.3)}$$

References

[1] M.C. Campi and Erik Weyer, Identification with finitely many data points: The LSCR approach, *Semi-plenary presentation. In Proc. Symposium on System Identification*, SYSID 2006, New Castle, Australia, 2006.

[2] M.C. Campi and Erik Weyer, Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, Vol.41, pp1751-1764, 2005.

[3] M.C. Campi and Erik Weyer, Estimation of confidence regions for parameter of ARMA models – guaranteed non-asymptotic result. *In Proc. 42st Conf. on Decision and Control*, Maui, Hawaii, USA, 2003.

[4] M.C. Campi and Erik Weyer. Non-asymptotic confidence region for model parameter in the presence of unmodelled dynamics. *In Proc. 46th Conf. on Decision and Control*, New Orleans, LA, USA, 2007.

[5] Gordon, L. *Completely separating groups in subsampling. Annals of statistics*, Vol. 2, pp. 572-578.

[6] Ljung, L, *System identification – Theory for the user*. Second Edition, 1999. Pages 302-303,510,558-559.

[7] S. Garatti, M.C. Campi, and S. Bittanti. Assessing the quality of identified models through the asymptotic theory – When is the result reliable? *Automatica*. Vol. 40, pp. 1319-1332, 2004.

[8] S. Garatti, M.C. Campi, and S. Bittanti. The asymptotic model quality assessment for instrumental variable identification revisited. *Syst. Contr. Lett*., 55:494-500, 2006.

[9] Andrzej Bargiela, Interval and ellipsoidal uncertainty models. *Granular Computing: an emerging paradigm*, Pages 23-57, 2001.

[10] Söderström, T. and P. Stoica (1988), *System Identification. Prentice Hall*.

[11] E. Weyer and M.C. Campi. Non-asymptotic confidence ellipsoids for the least squares estimate. *Automatica*, 38:1529-1547, 2002.

[12] M.C. Campi and E. Weyer. Finite sample properties of system identification methods. *IEEE Trans. on Automatic Control*, AC-47:1329-1334, 2002.

[13] Bradley Efron, Robert J. Tibshirani. *An Introduction to the Bootstrap*.

[14] Fredrik Tjärnström, Quality Estimation of Approximate Models. *Department of Electrical Engineering, Linköpings universitet*, SE-581 83 Linköpings, Sweden, Chapter 5.

[15] Kun-Lin Hsieh, Yan-Kwang Chen, Ching-Cheng Shen. Bootstrap confidence interval estimates of the bullwhip effect. *Simulation modeling Practice and Theory*, Volume 15, Issue 8, September 2007, Pages 908-917.

[16] System Identification Toolbox, Matlab. http://www.mathworks.com/access/helpdesk_r13/help/toolbox/ident/ch3tut69.html.