GLOBAL NON-ASYMPTOTIC CONFIDENCE SETS FOR GENERAL LINEAR MODELS

M.C. Campi [*,1] **Erik Weyer** [**,2]

[*] *Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy. Email: campi@ing.unibs.it*
[**] *CSSIP, Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. Email: e.weyer@ee.mu.oz.au*

Abstract: In this paper we consider the problem of constructing confidence sets for the parameters of general linear models. Based on subsampling techniques and building on earlier exact finite sample results due to Hartigan, we compute the exact probability that the true parameters belong to certain regions in the parameter space. By intersecting these regions, a confidence set containing the true parameters with guaranteed probability is obtained. All results hold rigorously true for any finite number of data points and no asymptotic theory is involved. Moreover, prior knowledge on the uncertainty affecting the data is reduced to a minimum. The approach is illustrated on a simulation example, showing that it delivers practically useful confidence sets with guaranteed probabilities.

Keywords: Confidence sets, Uncertainty evaluation, Finite sample results, General linear models, System identification.

1. INTRODUCTION

Models of dynamical systems are used in many fields of science and engineering. It is widely recognised that a model is of limited use if no quality tag is attached to it. A good technique or methodology for model uncertainty evaluation should meet the following two requirements: 1) It must hold under general conditions, and 2) It should provide a non-conservative evaluation of the system uncertainties.

Regarding the first item we note that restrictive assumptions on the noise (e.g. that it is Gaussian or bounded) means that the theory is no longer applicable to many real life systems. The second point is important because loose uncertainty evaluations generate conservativeness in the belief that the model is less reliable than it actually is. For example, a robust controller looses in performance as the level of uncertainty increases.

One point that needs to be kept in mind is that, in system identification (e.g. Ljung (1999)), one always uses a *finite* number of data points. Likewise, for the evaluation of model quality and construction of confidence sets one will only have a finite amount of data available. Thus, a sound uncertainty evaluation method must provide results valid when the number of data is finite, and, possibly, small.

[1] This work is partly supported by MIUR under the project "New methods for Identification and Adaptive Control for Industrial Systems".
[2] This research is partly supported by the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP) under the Cooperative Research Centre scheme funded by The Commonwealth Government

Quite often, uncertainty evaluations and confidence ellipsoids are derived based on the asymptotic theory of system identification. It is common experience that this theory - though applied heuristically with a finite number of data points - in many situations delivers sensible results. On the other hand, the correctness of the results is not guaranteed, and contributions (e.g. Garatti et al. (2004)) have appeared that show that the asymptotic theory fails to be reliable in certain situations. Moreover, when the available data is scarce, using asymptotic results makes no sense. Thus, there is a need for developing techniques that provide results guaranteed for finite data samples.

In this paper we develop a methodology, 'Leave-out Sign-dominant Correlation Regions' (LSCR), for construction of confidence sets for a general linear system based on a finite number of data points. The constructed confidence sets have guaranteed probability of containing the true parameter. The developed theory is rigorously valid for any finite data sample and moreover, it provides non-conservative evaluation of the model uncertainties.

LSCR works well in cases where the asymptotic theory fails. Asymptotic confidence ellipsoids are based on a local Taylor expansion around the estimated parameter and can give misleading results when the estimate gets trapped in a local minimum. LSCR will in those cases typically deliver a confidence set which consists of disconnected regions around the true parameter and the local minima which have a predictive performance comparable to the true system parameter (see the simulation example in Section 4).

The mathematical approach in this paper is inspired by Hartigan (1969). The present paper departs from Hartigan (1969) in that we consider more general random sequences which allows us to deal with *dynamical* systems. Yet, the main underlying idea is still within Hartigan's framework.

Our earlier finite sample results (e.g. Campi and Weyer (2002) and Weyer and Campi (2002)) were data independent, in the sense that they were uniform with respect to the considered class of data generating systems, and they could essentially be evaluated without any data. Because of the uniformity, it was realised that the results could be quite conservative for the particular system at hand. The approach presented here is data based and uses data generated by the actual system at hand, and hence avoids the problems due to uniformity.

LSCR does not deliver a nominal model. Instead, it delivers a set of possible models to which the true system belongs with a guaranteed probability. In this respect the methodology has a lot in common with set membership identification, see e.g. Milanese and Vicino (1991), Bai et al. (1996), Giarre' et al. (1997). However, unlike the typical setting in set membership identification we do not need to assume that the disturbances are deterministic or bounded. Loosely speaking, one could view the developed methodology as a stochastic set membership approach to system identification, where the setting we consider is the standard stochastic setting for system identification from e.g. Ljung (1999) or Söderström and Stoica (1988), but where the outcomes are more in line with the outcomes from set membership identification.

The paper is organised as follows. In section 2, we construct regions which are such that we can calculate the exact probability that the true parameter belongs to them. In section 3 the confidence sets are obtained by taking the intersection of a number of the constructed regions. A simulation example is given in section 4 demonstrating that the method is practically useful.

2. CONFIDENCE REGIONS FOR LINEAR SYSTEMS

2.1 Data generating system

The data are generated by a general linear system

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

where $G^0(z^{-1})$ and $H^0(z^{-1})$ are stable rational transfer functions. z^{-1} is the backward shift operator ($z^{-1}y_t = y_{t-1}$). $H^0(z^{-1})$ is monic and has a stable inverse and $G^0(z^{-1})$ has a delay of 1 or more time units. $\{w_t\}$ is a zero-mean independent sequence (noise). No a-priori knowledge of the noise level is assumed. The system operates in open loop, that is $\{w_t\}$ and $\{u_t\}$ are independent. Closed loop systems are discussed in section 3.2.

2.2 Model structure

The model class consists of full order models

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

which are parameterised by θ. We assume that there exists a unique parameter θ^0 such that $G(z^{-1}, \theta^0) = G^0(z^{-1})$ and $H(z^{-1}, \theta^0) = H^0(z^{-1})$. Moreover, we assume that θ is restricted to a set Θ such that $H(z^{-1}, \theta)$ is monic, $G(z^{-1}, \theta)$, $H(z^{-1}, \theta)$ and $H^{-1}(z^{-1}, \theta)$ are stable and $G(z^{-1}, \theta)$ has a delay of 1 or more time units for all $\theta \in \Theta$.

Remark 2.1. It should be noted that the goal of the present paper is to construct confidence regions for the system parameter (as opposed to identifying a nominal model). Similar to other existing techniques for model quality evaluation (e.g. bootstrap techniques (Tjärnström and Ljung (2002))), a full description of the system is adopted. This does not in any way enforce a full order nominal model: one can use a reduced order nominal model and then verify its reliability using a full order model for quality evaluation.

2.3 Construction of confidence regions

We start by describing procedures for the determination of sets Θ_r^ϵ and Θ_s^u for which we can exactly calculate the probabilities $Pr\{\theta^0 \in \Theta_r^\epsilon\}$ and $Pr\{\theta^0 \in \Theta_s^u\}$. Confidence sets $\hat{\Theta}$ for θ^0 can then be constructed by taking the intersection of a number of the Θ_r^ϵ and Θ_s^u sets.

Procedure for the construction of $\Theta_\mathbf{r}^\epsilon$

(1) Compute the prediction errors

$$\epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = H^{-1}(z^{-1},\theta)y_t - H^{-1}(z^{-1},\theta)G(z^{-1},\theta)u_t \quad (2)$$

where t takes on a finite number of values, say, $1,\ldots,K$;

(2) Select an $r \geq 1$. For $t = 1+r,\ldots,N+r=K$, compute

$$f_{t-r,r}^\epsilon(\theta) = \epsilon_{t-r}(\theta)\epsilon_t(\theta);$$

(3) Let $I = \{1,\ldots,N\}$ and consider a collection G of subsets $I_i \subseteq I$, $i = 1,\ldots,M$, forming a group under the symmetric difference operation (i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G$, if $I_i, I_j \in G$). Compute

$$g_{i,r}^\epsilon(\theta) = \sum_{k\in I_i} f_{k,r}^\epsilon(\theta), \quad i = 1,\ldots,M; \quad (3)$$

(4) Select an integer q in the interval $[1,(M+1)/2)$ and find the region Θ_r^ϵ such that at least q of the $g_{i,r}^\epsilon(\theta)$ functions are bigger than zero and at least q are smaller than zero.

The intuitive idea behind this construction is that, for $\theta = \theta^0$, the functions $g_{i,r}^\epsilon(\theta)$ assume positive or negative value at random ($\epsilon_t(\theta^0)$ is a zero mean independent sequence), so it is unlikely that almost all of them are positive or that almost all of them are negative. Since point 4 in the construction of Θ_r^ϵ discards regions where all $g_{i,r}^\epsilon(\theta)$'s but a small fraction (q should be taken to be small compared to M, see Theorem 2.1 below) are of the same sign, we expect that $\theta^0 \in \Theta_r^\epsilon$ with high probability. This is put on solid mathematical grounds in Theorem 2.1 below.

The procedure for construction of the sets Θ_s^u is in the same spirit. The only difference being that the empirical autocorrelations in point 2 are replaced by empirical cross correlations between the input signal and the prediction error.

Procedure for the construction of $\Theta_\mathbf{s}^\mathbf{u}$

(1) Compute the prediction errors as in equation (2).

(2) Select an integer $s \geq 1$. For $t = 1+s,\ldots,N+s=K$, compute

$$f_{t-s,s}^u(\theta) = u_{t-s}\epsilon_t(\theta);$$

(3) Let $I = \{1,\ldots,N\}$ and consider a collection G of subsets $I_i \subseteq I$, $i = 1,\ldots,M$, forming a group under the symmetric difference operation. Compute

$$g_{i,s}^u(\theta) = \sum_{k\in I_i} f_{k,s}^u(\theta), \quad i = 1,\ldots,M;$$

(4) Select an integer q in the interval $[1,(M+1)/2)$ and find the region Θ_s^u such that at least q among functions $g_{i,s}^u(\theta)$ are bigger than 0 and at least q are smaller than zero.

The next theorem gives the exact probability that the true parameter θ^0 belongs to one particular of the above constructed sets.

Theorem 2.1. Assume that variables w_t admit a density (so that $Pr\{w_t = c\} = 0$, for any t and any real c) and that they are symmetrically distributed around zero. Furthermore assume that $Pr\{u_t = 0\} = 0$, for all t. Then, the sets Θ_r^ϵ and Θ_s^u constructed above are such that:

$$Pr\{\theta^0 \in \Theta_r^\epsilon\} = 1 - 2q/M, \quad (4)$$

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

Proof. See Campi and Weyer (2003,2004). ∎

Remark 2.2. The noise assumption is mild enough to accommodate a number of situations. In particular, one can describe possible outliers by allowing the noise to take on large values with small probability. Importantly, the procedures return regions of guaranteed probability despite we do not assume any a-priori knowledge on the noise level: the noise level enters the procedures through data only. ∎

Theorem 2.1 quantifies the probability that θ^0 belongs to the regions Θ_r^ϵ and Θ_s^u. It holds for any finite N and introduces no conservativeness at all since $1-2q/M$ is the exact probability, not a lower bound of it. Each one of the sets Θ_r^ϵ and Θ_s^u is a non-asymptotic confidence set in its own right. However, each one of these sets will usually be unbounded in some directions of the parameter space. A practically useful confidence set $\hat{\Theta}$ can be obtained by intersecting a number of the sets Θ_r^ϵ and Θ_s^u, i.e.

$$\hat{\Theta} = \cap_{r=1}^{n_\epsilon}\Theta_r^\epsilon \cap_{s=1}^{n_u} \Theta_s^u. \quad (6)$$

The next obvious question is how to choose n_ϵ and n_u in order to obtain confidence sets that are bounded and concentrated around the true parameter θ^0. This question will be dealt with in the next section. We conclude this section with a fact which is immediate from Theorem 2.1.

Theorem 2.2. Under the assumptions of Theorem 2.1,

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

where $\hat{\Theta}$ is given by (6). ∎

The inequality in the theorem is due to that the sets $\{\theta^0 \notin \Theta_r^\epsilon\}$, $\{\theta^0 \notin \Theta_s^u\}$, $r = 1, \ldots, n_\epsilon$, $s = 1, \ldots, n_u$ may be overlapping.

3. CONFIDENCE SETS FOR DIFFERENT MODEL CLASSES

A good evaluation method must have two properties: the provided region must have guaranteed probability (and this is what Theorems 2.1 and 2.2 deliver); and the region must be bounded, and, in particular, it should concentrate around θ^0 as the number of data points increases. How to choose n_ϵ and n_u in in order to achieve this for ARMA and ARMAX models was treated in Campi and Weyer (2003, 2004). Here we provide guidelines for a general linear model class.

3.1 General linear models

A first observation is that each time we consider a set Θ_r^ϵ or Θ_s^u we exclude a region in the parameter space where θ^0 is unlikely to belong. By intersecting some of these sets without paying too much attention to any theory a-priori guaranteeing that the obtained region is bounded around the true parameter, it is likely (and it is experienced in simulation examples) that we come up with a satisfying confidence set. Here we develop a theoretical argument to substantiate the fact that using correlations permits to spot the true θ^0

Let us consider an infinite number of correlations: $E[u_{t-s}\epsilon_t(\theta)] = 0$ for all $s \geq 1$ and $E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)] = 0$ for all $r \geq 1$. Next we show that the only parameter satisfying them all is θ^0 (if u_t is sufficiently exciting).

The prediction error is given by

$$\begin{aligned}\epsilon(t,\theta) =& H^{-1}(z^{-1},\theta)(G^0(z^{-1}) - G(z^{-1},\theta))u_t + \\ & H^{-1}(z^{-1},\theta)H^0(z^{-1})w_t,\end{aligned}$$

and the conditions $E[u_{t-s}\epsilon_t(\theta)] = 0$ for all $s \geq 1$ becomes:

$$\int_{-\pi}^{\pi} H^{-1}(e^{-i\omega},\theta)\tilde{G}(e^{-i\omega},\theta)\Phi_u(\omega)e^{iws}d\omega = 0,$$

for all $s \geq 1$, where $\tilde{G}(e^{-i\omega},\theta) = G^0(e^{-i\omega}) - G(e^{-i\omega},\theta)$. This gives $G^0(e^{-i\omega}) = G(e^{-i\omega},\theta)$ provided that u is sufficiently exciting. Using this we also find that the conditions $E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)] = 0$ for all $r \geq 1$ can be written as

$$\int_{-\pi}^{\pi} |H^{-1}(e^{-i\omega},\theta)H^0(e^{-i\omega})|^2\lambda_w^2 e^{iwr}d\omega = 0$$

for all $r \geq 1$. It follows that $H^{-1}(e^{-i\omega},\theta)H^0(e^{-i\omega})$ is constant, and since $H(z^{-1},\theta)$ and $H^0(z^{-1})$ are monic, it follows that $H(e^{-i\omega},\theta) = H^0(e^{-i\omega})$.

In the construction of $\Theta_{i,r}^\epsilon$ and $\Theta_{i,s}^u$, we compute the sample correlations $g_{i,r}^\epsilon$ and $g_{i,s}^u$. As $N \to \infty$, the functions $\frac{1}{N_i}g_{i,r}^\epsilon(\theta) \to E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)]$ and $\frac{1}{N_i}g_{i,s}^u(\theta) \to E[u_{t-s}(\theta)\epsilon_t(\theta)]$, provided that the number of elements in each set I_i also tends to infinity. It is easy to construct groups with this property, see Gordon (1974). The normalisation $1/N_i$ is immaterial in the construction of the confidence regions, since all that matters is whether $g_{i,r}^\epsilon(\theta)$ and $g_{i,s}^u(\theta)$ are smaller or greater than zero.

This means that, if we could compute an infinite number of correlations, we would, under some additional uniformity conditions, end up with a confidence set which concentrates around the true parameter as the number of data points increases. Computing an infinite number of correlations is of course not feasible, and moreover the guaranteed probability $1 - (n_\epsilon + n_u)2q/M$ goes to zero as $n_\epsilon, n_u \to \infty$. However, in practice one only needs to compute a finite number of empirical correlations, and we next suggest a heuristic guideline for computing the correlations.

Guideline 1. Compute as many empirical correlations as there are parameters in the model. Let n be the number of parameters which $G(z^{-1},\theta)$ and $H(z^{-1},\theta)$ have in common, m the number of parameters which appear exclusively in $G(z^{-1},\theta)$ and p the number of parameters which appear exclusively in $H(z^{-1},\theta)$. Choose at least m correlations $u_{t-s}\epsilon_t(\theta),\ s = 1, \ldots, m$, and at least p correlations $\epsilon_{t-r}(\theta)\epsilon_t(\theta),\ r = 1, \ldots, p$. When choosing the last n correlations take into account the a priori information about the energy in the signals u_t and w_t and how exciting they are. Favour correlations of the type $u_{t-s}\epsilon_t(\theta)$ if u_t is the stronger signal and correlations of the type $\epsilon_{t-r}(\theta)\epsilon_t(\theta)$ if w_t is the stronger signal. ∎

Note that it is only the shape of the obtained set which is affected by the guideline. Theorem 2.1 and 2.2 are valid so the confidence set has a guaranteed probability.

3.2 Closed loop systems

Again consider a general linear system (1), but assume now that the input is generated by a feedback controller

$$u_t = K(z^{-1})(\tilde{r}_t - y_t)$$

where $\tilde{r}_t$ is a reference signal, so that the closed loop system is stable. We have that

$$\begin{aligned}u_t &= (1 + KG^0)^{-1}K\tilde{r}_t - (1 + KG^0)^{-1}KH^0w_t \\ y_t &= (1 + KG^0)^{-1}KG^0\tilde{r}_t + (1 + KG^0)^{-1}H^0w_t.\end{aligned}$$

In this context, we consider the closed loop system with inputs $\tilde{r}_t$ and w_t and compute $\hat{y}_t(\theta)$ from the $\tilde{r}_t$ and y_t signals. The corresponding prediction error is:

$$\begin{aligned}\epsilon(t,\theta) =& (1 + KG^0)^{-1}H^{-1}K(G^0 - G)\tilde{r}_t + \\ & (1 + (1 + KG^0)^{-1}(G - G^0)K)H^{-1}H^0w_t.\end{aligned}$$

Then, by imposing correlation relations $\tilde{r}_{t-s}\epsilon_t(\theta)$ and $\epsilon_{t-r}(\theta)\epsilon_t(\theta)$, nothing change in the probability analysis of section 2 and Theorems 2.1 and 2.2 remain valid.

As in section 3.1, by imposing $E[\tilde{r}_{t-s}\epsilon_t(\theta)] = 0, \ \forall s \geq 1$, it follows that $G = G^0$ provided $\tilde{r}_t$ is sufficiently exciting, and imposing $E[\epsilon_{t-r}(\theta)\epsilon_t(\theta)] = 0, \ \forall r \geq 1$ gives $H = H^0$. Therefore, considerations similar to those in section 3.1 for the choice of the correlation functions apply in this closed loop context as well.

4. SIMULATION EXAMPLE

The following example is taken from Garatti et al. (2004). Consider the system

$$y_t = \frac{b^0 z^{-1}}{1 + a^0 z^{-1}} u_t + (1 + h^0 z^{-1}) w_t, \qquad (7)$$

with $\theta^0 = [a^0 \ b^0 \ h^0]^T = [-0.7 \ 0.3 \ 0.5]^T$. $\{w_t\}$ is white Gaussian noise with variance 1, and the input is generated by

$$u_t = \tilde{r}_t - y_t \qquad (8)$$

where $\tilde{r}_t$ is white Gaussian noise with variance 10^{-6}.

In Garatti et al. (2004) it is shown that for this system the asymptotic variance expressions from system identification theory can give misleading results. Using a standard quadratic criterion $V_N(\theta) = 1/N \sum_{t=1}^{N} \epsilon^2(t, \theta)$, the estimate is given by $\hat{\theta}_N = \arg\min_\theta V_N(\theta)$. Asymptotically $\hat{\theta}_N$ converges to a value which minimises $V(\theta) = E\epsilon^2(t, \theta)$. For $\tilde{r}_t \equiv 0$, there are two isolated parameters which minimise $V(\theta)$. These values are the true parameter θ^0 and $\theta^* = [h^0 \quad a^0 - h^0 + b^0 \quad a^0]^T$. When the input signal is different from zero, but poorly exciting, the only minimum is θ^0, but $V(\theta^0)$ and $V(\theta^*)$ are close, and since the estimate is found by minimising $V_N(\theta)$, it will often end up being close to θ^* which is now only a local minimum.

The asymptotic theory for evaluation of the variance of $\hat{\theta}_N - \theta^0$ is based on a Taylor series expansion of $\sqrt{N} V_N'(\theta)$ around the true parameter θ^0 ($'$ and $''$ denote first and second derivative w.r.t. θ)

$$\begin{aligned} 0 &= \sqrt{N} V_N'(\hat{\theta}_N) \\ &= \sqrt{N} V_N'(\theta^0) + V_N''(\xi_N) \sqrt{N} (\hat{\theta}_N - \theta^0) \end{aligned}$$

where ξ_N is a point between $\hat{\theta}_N$ and θ^0. When the asymptotic expressions are used in the finite sample case $V''(\xi_N)$, is replaced by $V_N''(\hat{\theta}_N)$. This can give rise to a large error when $\hat{\theta}_N$ is far from θ^0 and this is what happens in this example. The net result is that the obtained confidence region is deceivingly small

As is clear from above, one reason why the asymptotic theory gives unreliable results is that it is local in nature in the sense that it is based on a Taylor expansion. It will only deliver a confidence set around the estimated parameter. This is in contrast to the non-asymptotic theory developed here which is global in nature as no local approximations are involved.

Returning to our approach for generating a confidence region we consider a full order model $y_t = \frac{bz^{-1}}{1+az^{-1}} u_t + (1 + hz^{-1}) w_t$. The prediction errors are given by

$$\epsilon_t(\theta) = \frac{1}{1 + hz^{-1}} y_t - \frac{bz^{-1}}{(1 + az^{-1})(1 + hz^{-1})} u_t = \frac{1 + (a + b)z^{-1}}{(1 + az^{-1})(1 + hz^{-1})} y_t - \frac{bz^{-1}}{(1 + az^{-1})(1 + hz^{-1})} \tilde{r}_t$$

As the system operates in closed loop, we consider $\tilde{r}_t$ as the input signal, and the model structure is $y_t = G(z^{-1}, \theta)\tilde{r}_t + H(z^{-1}, \theta) w_t$ with

$$G(z^{-1}, \theta) = \frac{bz^{-1}}{1 + (a + b)z^{-1}}$$

and

$$H(z^{-1}, \theta) = \frac{(1 + hz^{-1})(1 + az^{-1})}{1 + (a + b)z^{-1}}$$

We have three parameters, one which belongs to $H(z^{-1})$ only, and two which belong to both $G(z^{-1})$ and $H(z^{-1})$. Using the Guideline in section 3.1, we compute three correlations, one of them being $\epsilon_{t-1}(\theta)\epsilon_t(\theta)$. As $\tilde{r}_t$ is a poorly exciting signal compared to w_t, we choose the other two correlations to be $\epsilon_{t-2}(\theta)\epsilon_t(\theta)$ and $\epsilon_{t-3}(\theta)\epsilon_t(\theta)$.

We generated 2047+3 data points ($N = 2047$) according to (7) and (8). The group was constructed as in Gordon (1974) ($M = 2048$), and we computed

$$g_{i,1}^{\epsilon}(\theta) = \sum_{k \in I_i} \epsilon_k(\theta)\epsilon_{k-1}(\theta), \ \ r = 1, 2, 3$$

in the parameter space, making the standard assumptions that the open and closed loop systems were stable ($|a| < 1$, $|a+b| < 1$) and that $H(z^{-1}, \theta)$ has a stable inverse, ($|h| < 1$). We excluded the regions in the parameter space where 0 was among the 34 smallest or largest values of any of the three correlations above to obtain a $1 - 3 \cdot 2 \cdot 34/2048 = 0.9004$ confidence set. The confidence set is shown in Figure 1. The set consists of two separate regions, one around the true parameter θ^0 and one around θ^* the other minimum of $V(\theta)$ when $\tilde{r}_t \equiv 0$. This illustrates the global features of the approach, producing two separate regions far apart in the parameter space as the confidence set.

The parameter estimate itself is in this case close to θ^* and the asymptotic 90% confidence ellipsoid is shown in Figure 2 together with the part of our non-asymptotic confidence set which is concentrated around θ^*. As we can see, the asymptotic theory, due to its local nature, produces a misleading result, since the confidence region is situated around a parameter value corresponding to a local minimum and it does not include the true parameter θ^0. A close up of the part of the non-asymptotic confidence region around the true parameter θ_0 is shown in Figure 3.

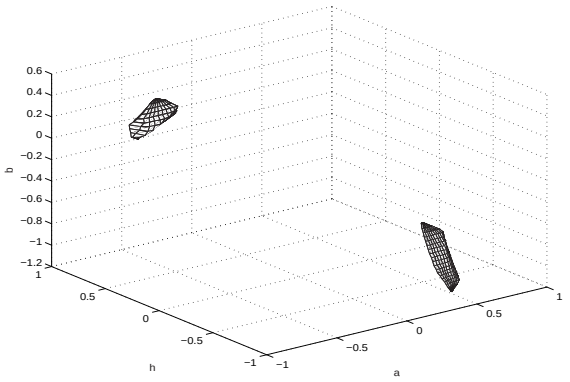

Fig. 1. 90% confidence region.

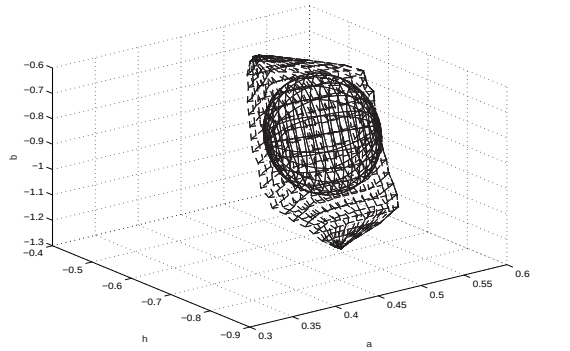

Fig. 2. Asymptotic confidence 90% ellipsoid (-), and the part of the non-asymptotic confidence set around θ^* (- -).

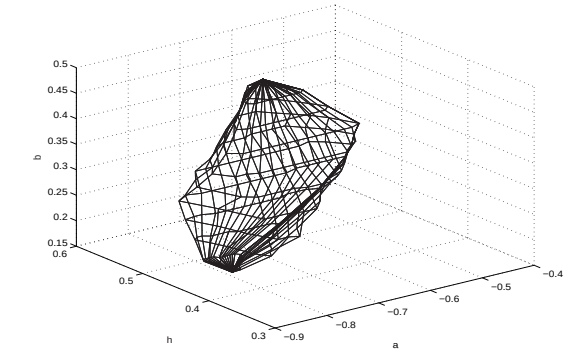

Fig. 3. Close up of the non-asymptotic confidence region around θ^0.

5. CONCLUSIONS

In this paper, we have derived a method called 'Leave-out Sign-dominant Correlation Regions' (LSCR) for the construction of confidence sets for general linear models. As the name indicates, LSCR is based on computing empirical correlation functions using subsamples and discarding regions in the parameter space where only a small fraction of the empirical functions are greater or smaller than zero. LSCR is grounded on a solid theoretical basis, giving guaranteed probabilities for the true parameter to belong to the constructed set for any finite number of data points. As illustrated by the simulation example, it produces practically useful confidence sets, and it is a global approach, delivering disconnected confidence regions when that is appropriate.

REFERENCES

Bai E.W., K.M. Nagpal and R. Tempo (1996). "Bounded-error parameter estimation: noise models and recursive algorithms." *Automatica*, Vol. 32., pp. 985-999.

Campi, M.C. and E. Weyer (2002). "Finite sample properties of system identification methods". *IEEE Trans. on Automatic Control*, Vol. AC-47, pp. 1329-1334.

Campi, M.C. and E. Weyer (2003). "Estimation of confidence regions fro the parameters of ARMA models - guaranteed non-asymptotic results". *Proc of IEEE CDC03*, pp. 6009-6014, Maui, Hawaii, US, December 2003.

Campi, M.C. and E. Weyer (2004). "Non-asymptotic confidence sets for the parameters of ARMAX models". In proceedings of *ALCOSP 04*, Yokohama, Japan, September 2004.

Garatti S., S. Bittanti and M.C. Campi (2004). "Assessing the quality of identified models through the asymptotic theory - When is the result reliable?". *Automatica*, Vol. 40., no. 8, pp. 1319-1332.

Giarre' L., B.Z. Kacewicz and M. Milanese (1997). "Model quality evaluation in set membership identification." *Automatica*, Vol. 33., no. 6, pp. 1133-1139.

Gordon L. (1974). "Completely separating groups in subsampling". *Annals of Statistics*, Vol. 2, pp. 572-578.

Hartigan J. A. (1969). "Using subsample values as typical values". *Journal of American Statistical Association*, Vol. 64, pp. 1303-1317.

Ljung, L. (1999). *System Identification - Theory for the User.* 2nd Ed. Prentice Hall.

Milanese M. and A. Vicino (1991). "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview." *Automatica*, Vol. 27., no. 6, pp. 997-1009.

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

Tjärnström F., and L. Ljung (2002). "Using the bootstrap to estimate the variance in the case of undermodelling" *IEEE Transactions on Automatic Control*. Vol. 47, no. 2, pp. 395-398.

Weyer E. and M.C. Campi (2002). "Non-asymptotic confidence ellipsoids for the least squares estimate." *Automatica*, Vol. 38., no. 9, pp. 1539-1547.