Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

Non-asymptotic confidence sets for input-output transfer functions

Marco C. Campi and Erik Weyer

Abstract— **In this paper we consider the problem of constructing confidence sets for the parameters of linear systems in the presence of arbitrary noise. The developed LSCR method (Leave-out Sign dominated Correlation Regions) delivers confidence regions for the model parameters with guaranteed probability. 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 even when the noise is biased.**

I. Introduction

Models of dynamical systems are used in many fields of science and engineering. It is however widely recognised that a model is of limited value if no quality tag which tells us the accuracy of the model, is supplied. It is desirable that the method used for assessment of the model accuracy delivers a non-conservative evaluation of the system uncertainties under general conditions. Moreover, since there will always only be a finite amount of data available for evaluation of model uncertainties, the uncertainty description must be valid when the number of data points is finite.

In this paper we present a procedure which gives guaranteed non-asymptotic confidence regions for the model parameters of a dynamical system subjected to arbitrary noise. The procedure consists of a simple input design step, followed by an algorithm named LSCR (Leave-out Sign-dominant Correlation Regions) which construct the confidence set from the observed data. The procedure provides a non-conservative evaluation of the uncertainties in the model parameters, and it is valid for any finite number of data points.

Roughly speaking the problem considered is as follows:

An unknown dynamical system formula

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

as in Figure 1 is given. The transfer function $G^0(z^{-1})$ belongs to a set of transfer functions $G(\theta, z^{-1})$ parameterized by θ, that is $G^0(z^{-1}) = G(\theta^0, z^{-1})$ for some θ^0. The structure of the model class $G(\theta, z^{-1})$ is known, but θ^0 itself is unknown. The noise w_t describes all other source of variation in y_t apart from those caused by u_t, and w_t is independent of u_t. Our task is to design an input signal u_t, $t = 1, \ldots, N$, and a procedure which yields a confidence region $\hat{\Theta}_N$ for θ^0 with the property that

$$Pr\{\theta^0 \in \hat{\Theta}_N\} = 1 - \delta$$

where the number of data points N, and the confidence level δ, can be arbitrary chosen.

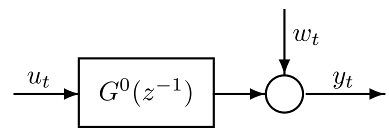

Fig. 1. Dynamical system.

Note that the procedure must work for any noise w_t. The noise can be large or small, correlated or uncorrelated, zero mean or even biased. Moreover, no a-priori information on the noise which can be used by the procedure, is available. This is important since noise characteristics are hardly known in practice. w_t can e.g. describe external influences generated by other systems, measurement noise, etc.

The philosophy behind the problem formulation is that we let the data speak, without assuming what they have to tell us through a-priori assumptions on noise. The size of the obtained $\hat{\Theta}_N$ will of course depend on the noise, but it must be the best possible given the present noise level. Moreover, the results must hold for any given finite number N of data points, since in practice the number of data points is finite, and possibly small.

The ideas of the proposed approach are illustrated in the next preview example. Then, subsection I-B discusses the contribution of the present paper relative to the existing literature. The detailed developments are presented in the following sections.

A. A preview example

Here the main ideas are illustrated. Consider the system

$$y_t = b^0 u_t + w_t \quad (1)$$

where $b^0 = 1$ and $\{w_t\}$ is an independent sequence of normally distributed variables with mean 0.5 and variance 1, i.e. the noise is biased. The distribution of w_t is given for completeness, but it is not used in the algorithm.

Our goal is to generate $N = 15$ input data $u_1, \ldots, u_{15}$ which enable us to construct a confidence interval $\hat{\Theta}_{15}$ such that $Pr\{b^0 \in \Theta_{15}\} = 0.75$.

M.C. Campi is with the Department of Electronic for Automation, University of Brescia, via Branze 38, 25123, Brescia, Italy. campi@ing.unibs.it

E. Weyer is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. e.weyer@ee.unimelb.edu.au

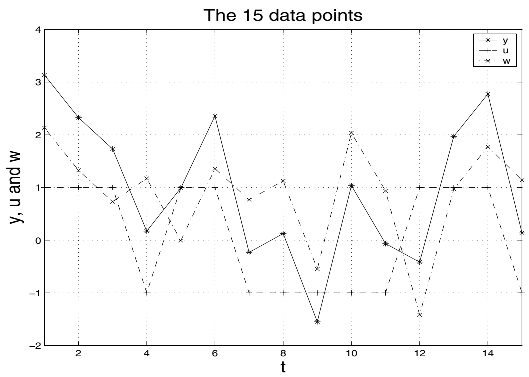

Fig. 2. The input, output and noise sequence used in the preview example.

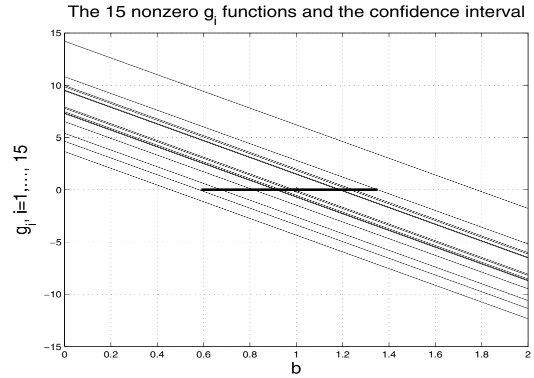

Fig. 3. The $g_i(b)$ functions for the preview example together with the confidence interval.

Input design. Let $u_t,\ t = 1, \ldots, 15$, be independent and identically distributed (iid) with distribution

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

(In fact any input signal which is iid and symmetrically distributed around 0 will do.) 15 data points were generated according to (1) and (2) and shown in Figure 2.

Procedure for construction of a confidence interval $\hat{\Theta}_N = [\hat{\theta}_{\min,N}, \hat{\theta}_{\max,N}]$.

Rewrite the system as a model with generic parameter b:

$$y_t = bu_t + w_t$$

The predictor and prediction error associated with the model are

$$\hat{y}_t(b) = bu_t, \quad \epsilon_t(b) = y_t - \hat{y}_t(b) = y_t - bu_t$$

Next we compute the prediction errors $\epsilon_t(b)$ for $t = 1, \ldots, 15$ and calculate

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

Using the $f_t(b)$'s, we want to form empirical estimates of the correlation $E[u_t\epsilon_t(b)]$. We note that $E[u_t\epsilon_t(b)] = (b^0 - b)$ which is equal to 0 for $b = b^0$ and different from zero for $b \neq b^0$. Hence, the empirical estimates will be zero mean random variables for $b = b^0$. Based on this observation, we compute a number of estimates of the correlation using different subsets of the data, and we discard those regions in parameter space where the empirical estimates take positive (or negative) value too many times. These empirical estimates, however, need to be carefully constructed as illustrated in the following.

First, we generate a set G of subsets of $I = \{1, \ldots, 15\}$ which is a group with respect to the symmetric difference, i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G$, if $I_i, I_j \in G$. The set I is the index set for the 15 functions $f_1(b), f_2(b), \ldots, f_{15}(b)$, and each set in the group G gives the indices of the functions $f_i(b)$ used for computing one particular empirical estimate. The group used in this example is given in Appendix A.

The estimates of the scaled correlation $E[u_t\epsilon_t(b)]$ (scaled, since we do not divide with the number of elements in I_i) are then given by

$$g_i(b) = \sum_{k \in I_i} f_k(b), \quad i = 1, \ldots, 16$$

($g_i(b) = 0$ if $I_i = \emptyset$). The 15 nonzero $g_i(b)$'s are plotted in Figure 3 as functions of b.

We recognise that it is very unlikely that all the $g_i(b^0)$'s have the same sign, and we therefore discard the rightmost and leftmost regions where at most one function out of the 15 non-zero functions is less than zero or greater than zero, hence the name of the method; Leave-out Sign-dominant Correlation Regions. The resulting interval $\hat{\Theta}_{15} = [0.59, 1.35]$, is the confidence region for b^0. It is a rigorous fact (stated in Theorem 3.1) that the confidence region constructed this way has probability $1 - 2 \cdot 2/16 = 0.75$ to contain the true parameter value b^0. Despite the fact that the noise is biased, the presented procedure provides a confidence interval around the true parameter value. The developed procedures stands on a solid theoretical footing and it delivers confidence regions with guaranteed probabilities which holds rigorously true for any finite number of data points and any arbitrary noise.

As expected, due to the small number of data points, this confidence interval is rather large and the associated probability is low. Next we repeated the experiment with 2047 data points using the group with incidence matrix $R(2047)$ (see Appendix A). We kept the region in parameter space where at least 50 of the 2047 nonzero $g_i(b)$ functions are greater than 0 and at least 50 are smaller than zero. The resulting interval $\hat{\Theta}_{2047} = [0.93, 1.06]$, contains the true parameter value b^0 with exact probability $1 - 50 \cdot 2/2048 = 0.9512 > 95\%$ (see Theorem 3.1).

B. Contribution and discussion of the existing literature

The results in this paper build on our previous work Campi and Weyer (2005). In that paper we assumed that the true transfer functions from both the noise and the input signal to the output signal belonged to the model class, and we derived non-asymptotic confidence sets for the model parameters of both transfer functions. These results are significantly strengthened in the present paper. The assumption that the true transfer function between the noise and the output belongs to the model class has been completely removed. In this contribution there are no assumptions on the noise, it can even be biased as seen in the preview example. Here we present a procedure for input design and an algorithm for constructing mathematically rigorous confidence sets for the system parameters valid for any number of data points regardless of what the noise is. As in Campi and Weyer (2005) the mathematical approach used in this paper is inspired by the work of Hartigan (1969).

A standard approach for deriving confidence sets for the parameters of dynamical systems is to employ asymptotic system identification theory (see e.g. Ljung (1999) or Söderström and Stoica (1988)). It is common experience of theorists and practitioners 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 may fail to be reliable in certain situations. Thus, there is a need for developing techniques that provide results guaranteed for finite data samples, and this is what the current paper delivers.

Similarly to set membership identification, the method presented in this paper returns regions for the true system parameter. However, unlike the setting in set membership identification, we do not make any assumptions about the disturbances. Further discussions on model quality evaluation and confidence sets for the parameters of dynamical systems are given in Campi and Weyer (2005).

C. Organisation of the paper

In this paper we present a basic version of our procedure applied to systems where the transfer function between input and output is given by $G^0(z^{-1}) = B^0(z^{-1})/A^0(z^{-1})$. In Section II we introduce the data generating system and state the assumptions on the system and the disturbance. Then in section III we present the LSCR algorithm for construction of confidence sets. We state results showing that the confidence sets have guaranteed probability for any finite number of data points and that they shrink around the true parameter as the number of data points increases. The procedure is illustrated on a simulation example in Section IV. Conclusions are given in Section V. The proofs are omitted due to space limitations, but will be provided in an upcoming publiction.

II. Data generating system

The data generating system is given by

$$y_t = G^0(z^{-1})u_t + w_t \tag{3}$$

where z^{-1} is the backward shift operator ($z^{-1}u_t = u_{t-1}$). Here $G^0(z^{-1}) = \frac{B^0(z^{-1})}{A^0(z^{-1})}$, with

$$\begin{aligned} A^0(z^{-1}) &= 1 + a_1^0 z^{-1} + \cdots + a_{n_a}^0 z^{-n_a} \\ B^0(z^{-1}) &= b_1^0 z^{-1} + \cdots + b_{n_b}^0 z^{-n_b} \end{aligned}$$

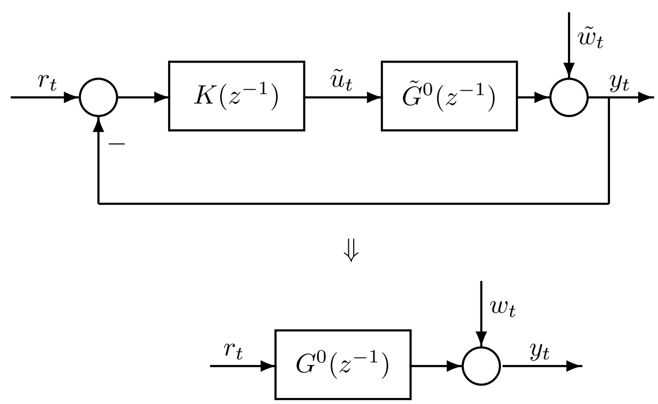

Fig. 4. Closed loop system recast as an open loop system.

Assumptions.

1) The user can choose the input signal u_t, and the choice of u_t does not affect w_t.
2) The model orders n_a and n_b are known and $A^0(z^{-1})$ and $B^0(z^{-1})$ are co-prime.

There are no assumptions on w_t. No upper bound on the magnitude is assumed, and it is allowed to have any non-zero mean and the auto correlation function can be arbitrary.

Remark 2.1: There is no loss of generality in having w_t additive at the output. Disturbances not entering at the output can for example be represented by $w_t = H^0(z^{-1})e_t$ where e_t is the real disturbance and $H^0(z^{-1})$ describes how e_t is filtered when it passes through the system. Since no assumption is made about w_t, there are no assumption on $H^0(z^{-1})$.

Remark 2.2: Closed loop systems can be cast in the present setting as shown in Figure 4 where r_t plays the role of u_t and

$$\begin{aligned} G^0(z^{-1}) &= \frac{\tilde{G}^0(z^{-1})K(z^{-1})}{1+\tilde{G}^0(z^{-1})K(z^{-1})} \\ w_t &= \frac{1}{1+\tilde{G}^0(z^{-1})K(z^{-1})}\tilde{w}_t \end{aligned}$$

The algorithm developed in this paper allows us to construct confidence sets for $\tilde{G}^0(z^{-1})$ without going via the closed loop transfer function $G^0(z^{-1})$ assuming that the controller $K(z^{-1})$ is known.

III. Algorithm for construction of confidence sets

The procedure consists of an easy experiment design step, followed by an algorithm for construction of the confidence sets based on the observed data. The confidence region is constructed as the intersection of a number of larger sets which are generated using the empirical correlation between the input signal and the prediction error as already illustrated in the preview example.

Experiment design.

Let $u_t,\ t = 1-n, \ldots, N$ be independent and symmetrically distributed around 0. $n = n_a + n_b$ is the number of system parameters.

Construction of confidence sets.

1) Compute the predictions[1]

$$\begin{aligned}\hat{y}_t(\theta) &= (1 - A(z^{-1}, \theta))y_t + B(z^{-1}, \theta)u_t \\ &= \phi_t^T \theta, \quad t = 1, \ldots, N\end{aligned}$$

where

$$\begin{aligned}A(z^{-1}, \theta) &= 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a} \\ B(z^{-1}, \theta) &= b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b} \\ \phi_t &= [-y_{t-1}, \ldots, -y_{t-n_a}, \\ &\quad u_{t-1}, \ldots, u_{t-n_b}]^T \\ \theta &= [a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}]^T\end{aligned}$$

2) Compute the prediction errors

$$\epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = y_t - \phi_t^T \theta$$

3) Form the vector ξ_t of Instrumental Variables (IV)

$$\xi_t = [u_{t-1}, \ldots u_{t-n}]^T, \quad t = 1, \ldots, N$$

and compute

$$f_t(\theta) = \xi_t \epsilon_t(\theta), \quad t = 1, \ldots, N$$

4) Let $I = \{1, \ldots, N\}$ and construct the group $G(N) = \{I_i,\ i = 1, \ldots, M\}$ of subsets of I given in Appendix A[2]. $G(N)$ is a group under the symmetric difference operation (i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G(N)$, if $I_i, I_j \in G(N)$). Compute[3]

$$g_i(\theta) = \sum_{t \in I_i} f_t(\theta) \quad i = 1, \ldots, M$$

5) Let $g_i^k(\theta)$ denote the kth element of $g_i(\theta)$ $k = 1, \ldots, n$. Select an integer q in the interval $[1, (M+1)/2n]$ Construct the regions $\hat{\Theta}_N^k$ such that at least q of the $g_i^k(\theta)$ functions are larger than 0 and at least q are smaller than 0. The confidence set is given by

$$\hat{\Theta}_N = \cap_{k=1}^n \hat{\Theta}_N^k \tag{4}$$

The intuitive idea is that for $\theta = \theta^0$ the functions $g_i^k(\theta) = \sum_{t \in I_i} u_{t-k}\epsilon_t(\theta)$ takes on positive and negative values at random since $g_i^k(\theta^0) = \sum_{t \in I_i} u_{t-k} A^0(z^{-1}) w_t$ and u_{t-k} is independent and symmetrically distributed around 0. It is therefore unlikely that only a small fraction of them are positive or negative, and point 5 in the algorithm excludes the regions in parameter space where this happens (q should be small compared to M). It is shown in Theorem 3.1 below that the probability that θ^0 belongs to $\hat{\Theta}_N^k$ is exactly $1 - 2q/M$.

The algorithm above has connections with Instrumental Variable (IV) methods for system identification. The main idea behind IV methods is that the prediction errors should be uncorrelated with past data, and the estimate is given by

$$\hat{\theta}_N = \left\{ \theta \,\middle|\, \sum_{t=1}^N \xi_t \epsilon_t(\theta) = 0 \right\}$$

where ξ_t is a vector made up of past data independent of the noise. In our approach the goal is not to come up with a particular value for the estimate, but to construct confidence sets, and we ensure that ξ_t and $\epsilon_t(\theta)$ are uncorrelated for $\theta = \theta^0$ by input design. The confidence set $\hat{\Theta}_N$ is constructed by excluding the regions in parameter space where any of the components of $\sum \xi_t \epsilon_t(\theta)$ takes on positive or negative values too many times when calculated on the subsets given by the group.

The next theorem gives the probabilities that θ^0 belongs to the constructed sets.

Theorem 3.1: Let u_t be independent an symmetrically distributed around 0, and let $I_i,\ i = 1, \ldots, M$, be the elements of the group used in point 4 in the algorithm. Moreover, assume that $Pr\{g_i(\theta^0) = 0\} = 0$ for all i for which $I_i \neq \emptyset$. Then, for any $k = 1, \ldots, n$,

$$Pr\{\theta^0 \in \hat{\Theta}_N^k\} = 1 - 2q/M \tag{5}$$

where $\hat{\Theta}_N^k$ is constructed in point 5 of the algorithm given above. ∎

An immediate consequence of Theorem 3.1 is

Corollary 3.2: Under the assumptions in Theorem 3.1,

$$Pr\{\theta^0 \in \hat{\Theta}_N\} \geq 1 - 2nq/M \tag{6}$$

where $\hat{\Theta}_N$ is given by (4). ∎

Note that the probability in (5) is the exact probability, not a lower bound. The inequality in (6) is due to that the events $\{\theta^0 \notin \hat{\Theta}_N^k\},\ k = 1, \ldots, n,$ may be overlapping. Also note that the stochastic element is the set $\hat{\Theta}_N$ since it is constructed from data, and that the probability is with respect to the input and the noise.

The only reason for the assumption $Pr\{g_i(\theta^0) = 0\} = 0$ is to prevent ties from occurring in point 5 of the procedure for constructing $\hat{\Theta}_N^k$, i.e. to prevent two functions to take on the value 0 on a set of non-zero measure in $\mathbb{R}^n$. The assumption is mild. It is for example satisfied whenever u_t and w_t admit densities. Moreover, the assumption can be completely removed by using a random ordering when several functions take on the value 0.

Theorem 3.1 and Corollary 3.2 shows that the constructed confidence sets have guaranteed probability. The next theorem shows that the confidence sets concentrate around the true parameter θ^0 as the number of data points increases.

Theorem 3.3: In addition to the assumptions of Theorem 3.1, assume that

[1]The predictors are obtained from (3) by ignoring w_t. The predictors are not the one step ahead predictors commonly used in system identification as these predictors require more knowledge about the noise than is available in the present setting.

[2]In fact any group with respect to symmetric differences can be used. A group is a non-empty set G with a binary operation $\circ$ such that: $a \circ b \in G,\ \forall a, b \in G$ and $a \circ (b \circ c) = (a \circ b) \circ c,\ \forall a, b, c \in G$; there exists an identity element $e \in G$ such that $a \circ e = e \circ a = a,\ \forall a \in G$; for every $a \in G$, there exists an inverse a^{-1} such that $a \circ a^{-1} = a^{-1} \circ a = e$.

[3]In order to get an estimate of the correlation functions we should divide by the number of elements in I_i. However, since only the sign of the g_i functions is important we have omitted this normalisation.

1) $A^0(z^{-1})$ is asymptotically stable.
2) $|u_t| \leq U$ for some U and $|w_t| \leq Kt^{\alpha}$ 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 ϵ neighbourhood of θ_0 for all $N > N(\epsilon)$. ∎

Remark 3.1: The assumption on w_t is very mild since w_t is allowed to grow unbounded. It is for example almost surely satisfied if w_t is a white Gaussian noise sequence. The assumption that u_t is mild too since u_t is a signal chosen by the user. Moreover, the assumption can be relaxed by assuming that the higher order moments are bounded in terms of the second order moment using Cramer's conditions, i.e. there exists a c such that $E|u_t|^k < c^{k-2}k!Eu_t^2$.

Corollary 3.2 and Theorem 3.3 show that the confidence sets have guaranteed probability for any finite number of data points, and moreover, as the number of data points increases the confidence sets shrink around the true parameter. Next we illustrate the method on a simulation example.

IV. Simulation example

The system is given by

$$y_t = a^0 y_{t-1} + b^0 u_{t-1} + v_t$$

where $a^0 = 0.7$ and $b^0 = 0.3$. u_t is independent and uniformly distributed on $[-\sqrt{3},\ \sqrt{3}]$, i.e. it is zero mean with variance 1. Here we have denoted the noise by v_t since w_t in the formulation (3) is given by $w_t = \frac{1}{1+a^0 z^{-1}} v_t$. For v_t we consider two different processes. The first one is lowpass filtered white noise

$$v_t = c_0 v_{t-1} + (1 - c_0^2) e_t$$

where $c_0 = 0.5$ and e_t is white Gaussian noise with variance 0.16. The second processes is biased lowpass filtered white noise

$$v_t^{\mathrm{b}} = v_t + 0.5$$

where v_t is given above. The predictor is given by

$$\hat{y}_t = a y_{t-1} + b u_{t-1}$$

For each noise scenario a data set of length $N = 1023$ was generated. The prediction errors are given by

$$\epsilon_t(a, b) = y_t - a y_{t-1} - b u_{t-1}, \quad t = 1, \ldots, 1023$$

and the group is constructed according to the procedure in Appendix A. This group has $M = 1024$ elements. Next, we computed the empirical correlations

$$\begin{aligned} g_{i,1}(a,b) &= \sum_{t \in I_i} u_{t-1}\epsilon_t(a,b), \quad i = 1, \ldots, 1024 \\ g_{i,2}(a,b) &= \sum_{t \in I_i} u_{t-2}\epsilon_t(a,b), \quad i = 1, \ldots, 1024 \end{aligned}$$

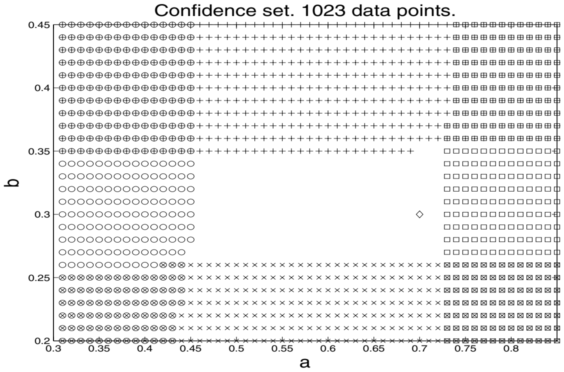

Fig. 5. Non-asymptotic confidence region for (a^0, b^0) (blank region). 1023 data points. w_t is lowpass filtered white noise. ◇ = true parameter.

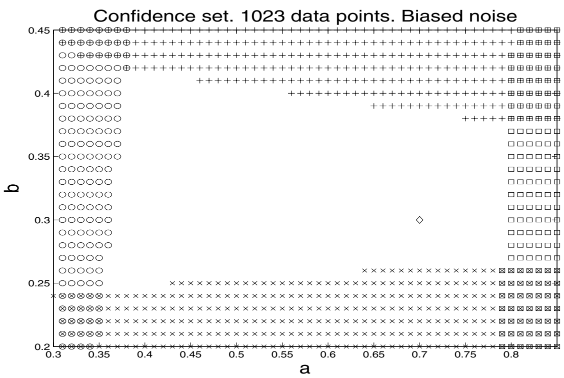

Fig. 6. Non-asymptotic confidence region for (a^0, b^0) (blank region). 1023 data points. w_t is biased lowpass filtered white noise. ◇ = true parameter.

In order to obtain a 95% confidence set we discarded those values of a and b for which zero was among the 12 largest or smallest values $g_{i,1}(a,b)$ and $g_{i,2}(a,b)$. Then according to Theorem 3.2 (a^0, b^0) belongs to the constructed region with probability at least $1 - 2 \cdot 2 \cdot 12/1024 = 0.9531$.

The obtained confidence sets are the blank areas in Figure 5 (unbiased noise) and 6 (biased noise). The areas marked with x is where 0 is among the 12 smallest values of $g_{i,1}$, the areas marked with + is where 0 is among the 12 largest values of $g_{i,1}$. Likewise for $g_{i,2}$ with the squares representing when 0 belongs to the 12 largest elements and the circles the 12 smallest. The true value (a^0, b^0) is marked with a diamond. As we can see, each step in the construction of the confidence sets excludes a particular region.

Using the algorithm for the construction of $\hat{\Theta}_N$ we have obtained a bounded confidence set with a guaranteed probability based on a finite number of data points. As no asymptotic theory is involved this is a rigorous finite sample result. Moreover, the results were obtained without using any a priori knowledge about the noise.

Next we increased the number of data points to $N = 4095$. The group from Appendix A has now $M = 4096$ elements, and in order to obtain a 95% confidence set, we now excluded the regions in parameter space where zero were among the

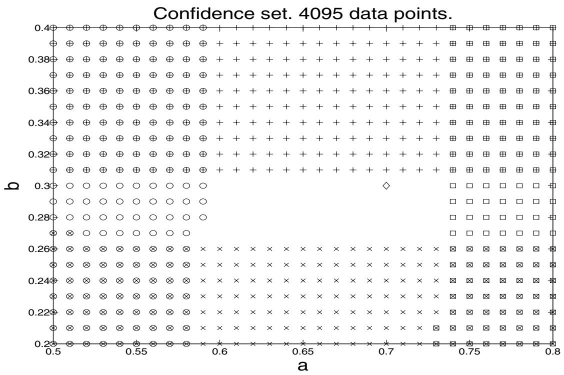

Fig. 7. Non-asymptotic confidence region for (a^0, b^0) (blank region). 4095 data points. w_t is lowpass filtered white noise. $\diamond$ = true parameter.

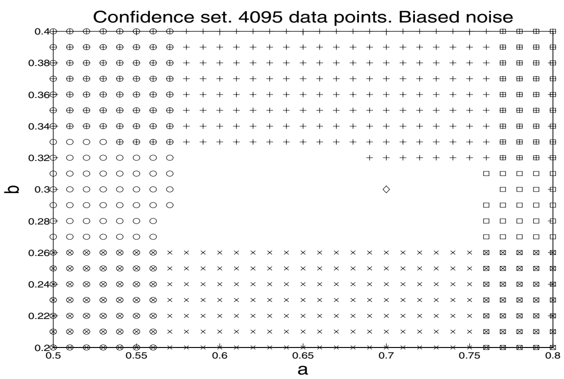

Fig. 8. Non-asymptotic confidence region for (a^0, b^0) (blank region). 4095 data points. w_t is biased lowpass filtered white noise. $\diamond$ = true parameter.

48 largest or smallest values of $g_{i,1}(a, b)$ and $g_{i,2}(a, b)$. The results are shown in Figure 7 and 8. The size of the sets are smaller than with 1023 data points illustrating that the confidence sets concentrates around the true parameters as the number of data points increases.

V. Conclusions

In this paper, we have extended the Leave-out Sign-dominated Correlation Regions (LSCR) method for the construction of confidence sets to systems with arbitrary noise. The method 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.

The developed methodology 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, and, as illustrated by the simulation examples, it produces practically useful confidence sets.

References

[1] Campi M.C. and E. Weyer (2005).“Guaranteed Non-asymptotic Confidence Regions in System Identification”. *Automatica*, Vol. 41, pp. 1751-1764.

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

[3] Gordon L. (1974). “Completely separating groups in subsampling”. *Annals of Statistics*, Vol. 2, pp. 572-578.

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

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

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

Appendix

A. Gordon’s construction of the incident matrix of a group

Given $I = \{1, \dots, N\}$, the incident matrix for a group $G(N) = \{I_i\}$ of subsets of I is a matrix whose (i, j) element is 1 if $j \in I_i$ and zero otherwise. In Gordon (1974), the following construction procedure for an incident matrix $\bar{R}$ is proposed where $I = \{1, \dots, 2^l - 1\}$ and the group has 2^l elements.

Let $R(1) = [1]$, and recursively compute $(k = 2, 3, \dots, l)$

$$R(2^k - 1) = \begin{bmatrix} R(2^{k-1} - 1) & R(2^{k-1} - 1) & 0 \\ R(2^{k-1} - 1) & J - R(2^{k-1} - 1) & e \\ 0^T & e^T & 1 \end{bmatrix}$$

where J and e are, respectively, a matrix and a vector of all ones, and 0 is a vector of all zeros. Then, let

$$\bar{R} = \begin{bmatrix} 0^T \\ R(2^l - 1) \end{bmatrix}$$

Gordon (1974) also gives construction of groups when the number of data points is different from $2^l - 1$.