Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics

Marco C. Campi, Sangho Ko, and Erik Weyer

Abstract— **This paper deals with the problem of constructing confidence regions for the parameters of truncated series expansion models. The models are represented using orthonormal basis functions, and we extend the "Leave-out Sign-dominant Correlation Regions" (LSCR) algorithm such that *non-asymptotic* confidence regions can be constructed in the presence of unmodelled dynamics. The constructed regions have guaranteed probability of containing the true parameters for any finite number of data points. The algorithm is first developed for FIR models and then generalized to orthonormal basis functions expansions. The usefulness of the developed approach is demonstrated for Laguerre models in a simulation example.**

I. INTRODUCTION

One of the intrinsic tasks in system identification is to evaluate how close the model is to the true system. This depends heavily on the quality and the size of the observed input-output data set and the specific rule used to construct a (set of) model(s) from the observed data.

This work focuses on truncated series expansion models represented by orthonormal basis functions and develops a method for constructing confidence regions for the coefficients of the series expansion using only finitely many input-output data points $\{u_k, y_k\}_{k=1,\cdots,N}$. For this purpose, we extend the LSCR (Leave-out Sign-dominant Correlation Regions) algorithm introduced in [1]. The algorithms in [1], [2] provide *non-asymptotic* confidence sets with a user-specified probability for the case where the true transfer function from the input signal to the output signal belongs to the model class. Here we remove the constraint that the true system must belong to the model class, and we consider truncated series models

$$G(z) = \sum_{k=1}^{L} \theta_k \mathcal{B}_k(z) \tag{1}$$

for the true system represented by an infinite series

$$G^0(z) = \sum_{k=1}^{\infty} \theta_k^0 \mathcal{B}_k(z) \tag{2}$$

using orthonormal basis functions $\{\mathcal{B}_k(z)\}$. Moreover we accommodate any noise sequence corrupting the output sequence.

Marco Campi is with the Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy. marco.campi@ing.unibs.it

Sangho Ko and Erik Weyer are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia. {s.ko,e.weyer}@ee.unimelb.edu.au

Typical examples of such basis functions are the pulse functions $\{z^{-k}\}$ corresponding to the FIR (Finite Impulse Response) models, the Laguerre models [3], the Kautz models [4], and more generally the orthonormal basis functions in [5] and [6].

The main novelty of the proposed approach as compared to the standard LSCR algorithm is the application of the *sign-function* in the computations of the correlation functions, and this allows us to deal with unmodelled dynamics.

In the next subsection we give a simple preview example which illustrates the main ideas of the proposed approach. Then in Section II the algorithm is presented at a general level for FIR models and extended to models represented by generalized orthonormal basis functions in Section III. A simulation example demonstrating the usefulness of the proposed approach is given in Section IV.

A. A preview example

To illustrate the main ideas of the paper, we present an introductory toy-example. Suppose that the true system is given by

$$y_t = \theta_0^0 u_t + \theta_1^0 u_{t-1} + n_t, \tag{3}$$

where $\theta_0^0 = 1$ and $\theta_1^0 = 0.1$, and the noise has been indicated with a generic n_t to signify that it can be arbitrary, and not just a white signal. Since the output y_t has weaker dependence on the past input u_{t-1} than on the current input u_t, we may want to find a non-dynamical link between u_t and y_t.

Our task is to generate 7 input data and to construct a guaranteed confidence interval for θ_0^0.

We first generate an input signal u_t, $t = 1, \cdots, 7$ which is independent and identically distributed (i.i.d.) with

$$u_t = \begin{cases} +1, & \text{with probability 0.5} \\ -1, & \text{with probability 0.5,} \end{cases} \tag{4}$$

and apply it to the system. The input-output data are shown in Fig. 1. We regard the term $\theta_1^0 u_{t-1}$ as unmodelled dynamics and construct a reduced-order predictor

$$\hat{y}_t(\theta) = \theta u_t. \tag{5}$$

The corresponding prediction error is given by

$$\epsilon_t(\theta) = y_t - \hat{y}_t(\theta) = y_t - \theta u_t. \tag{6}$$

We calculate

$$f_t(\theta) = \text{sign}[u_t \epsilon_t(\theta)], \quad t = 1, \cdots, 7, \tag{7}$$

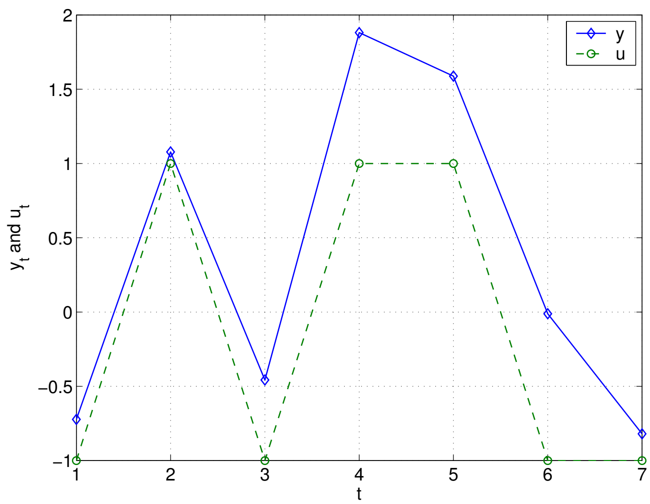

Fig. 1. Data for the preview example

where the sign-function is defined as

$$\mathrm{sign}[x] = \begin{cases} -1, & \text{for } x < 0, \\ 0, & \text{for } x = 0, \\ +1, & \text{for } x > 0. \end{cases}$$

Corresponding to the true parameter value, $\theta = \theta_0^0$, an easy inspection reveals that $\mathrm{sign}[u_t \epsilon_t(\theta_0^0)] = \mathrm{sign}[u_t(\theta_1^0 u_{t-1} + n_t)]$ is an independent and symmetrically distributed process. It is in fact a Bernoullian process taking on the values ± 1 with probability 0.5 each. Thus, based on this observation, we compute a number of scaled estimates of $E\{\mathrm{sign}[u_t \epsilon_t(\theta)]\}$ 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. The subsets are generated by forming a set $\mathbf{G}$ of subsets of $\mathbf{I} = \{1, \cdots, 7\}$ which is a group with respect to the symmetric difference, i.e., $(\mathbf{I}_i \cup \mathbf{I}_j) - (\mathbf{I}_i \cap \mathbf{I}_j) \in \mathbf{G}$, if $\mathbf{I}_i$, $\mathbf{I}_j \in \mathbf{G}$ (see [1]). The sets $\mathbf{I}_i$ in the group $\mathbf{G}$ gives the indices of the functions $f_t(\theta)$ used for computing one particular empirical estimate. The group considered in this example is described by the incident matrix below.

	1	2	3	4	5	6	7
$\mathbf{I}_0$	0	0	0	0	0	0	0
$\mathbf{I}_1$	1	1	0	1	1	0	0
$\mathbf{I}_2$	1	0	1	1	0	1	0
$\mathbf{I}_3$	0	1	1	0	1	1	0
$\mathbf{I}_4$	1	1	0	0	0	1	1
$\mathbf{I}_5$	1	0	1	0	1	0	1
$\mathbf{I}_6$	0	1	1	1	0	0	1
$\mathbf{I}_7$	0	0	0	1	1	1	1

Here each row corresponds to a subset. A 1 means that the element is in the set, while 0 means that the element is not in the set. The scaled empirical estimates are then given by

$$\bar{g}_i(\theta) = \sum_{t \in \mathbf{I}_i} f_t(\theta), \; i = 0, \cdots, 7 \tag{8}$$

($\bar{g}_0(\theta) = 0$ since $\bar{g}_i(\theta) = 0$, if $\mathbf{I}_i = \emptyset$). Since it is very unlikely that all the $\bar{g}_i(\theta^0)$'s have the same sign, we discard the regions in parameter space where all functions but at most one are less than the zero function $\bar{g}_0(\theta)$ or greater than the zero function $\bar{g}_0(\theta)$, hence the name of the method: Leave-out Sign-dominant Correlation Regions (LSCR). Since $f_t(\theta) = \mathrm{sign}[u_t \epsilon_t(\theta)]$ can take on only the values $-1, 1$ and 0, it is possible that two or more of the $\bar{g}_i(\theta)$ functions take on the same value on an interval. This tie can be broken by introducing a random ordering (e.g., by adding a random number ν_i, which is uniformly distributed between -0.2 and 0.2, to the $\bar{g}_i(\theta)$ functions except for $\bar{g}_0(\theta)$)

$$g_i(\theta) = \bar{g}_i(\theta) + \nu_i, \quad i = 1, \cdots, 7. \tag{9}$$

Next we plot $g_i(\theta)$, $i = 1, \cdots, 7$ as functions of θ and exclude the regions where at most one function is greater than zero and at most one is smaller than zero. The obtained $g_i(\theta)$ functions and the confidence interval are shown in Fig. 2. The confidence interval is $\hat{\mathbf{\Theta}} = [0.73 \; 1.07]$. It is a rigorous fact (stated in Theorem 1) that the confidence interval constructed this way has probability $1 - 2 \cdot 2/8 = 0.5$ of containing the true parameter value θ_0^0. In this example, the noise sequence n_t was a realization of a biased independent Gaussian process with mean 0.5 and variance 0.1. However, the noise characteristics are only provided for completeness and no knowledge about them was used in the algorithm.

Despite the facts that the system is not within the model set, the number of data points is small, and the noise is biased, the procedure still provides a rigorous confidence interval for the true parameter value.

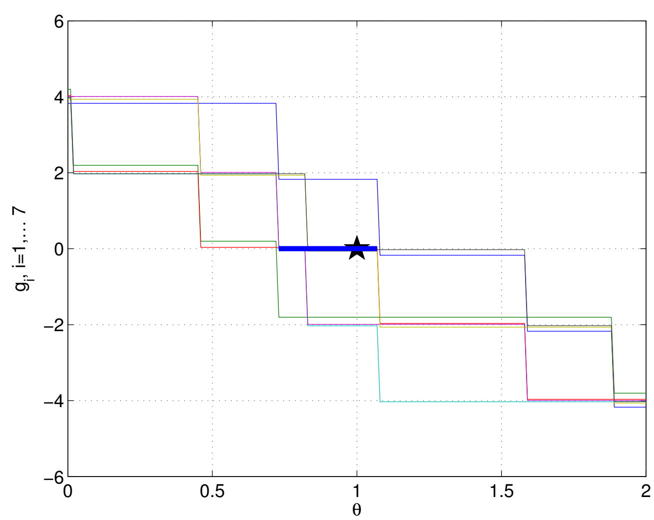

Fig. 2. The $g_i(\theta)$ functions for the preview example together with a 50% confidence interval (thick solid line) and the true parameter (★)

II. CONFIDENCE REGIONS WITH UNDERMODELLING

In this section we present the general algorithm for FIR models.

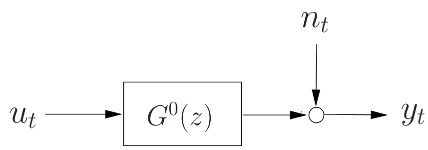

Fig. 3. The dynamical system

A. Problem definition

Data generating system:
Consider the following linear time-invariant stable discrete-time system with additive noise as shown in Fig. 3

$$y_t = G^0(z)u_t + n_t. \tag{10}$$

The transfer function $G^0(z)$ is represented by

$$G^0(z) = \sum_{k=1}^{\infty} \theta_k^0 z^{-k} \tag{11}$$

where $\{\theta_k^0\}_{k=1,2,\cdots}$ is the sequence of Markov parameters.

Assumption:

(**A1**) The noise sequence $\{n_t\}$ is independent of $\{u_t\}$, in the sense that our choice of u_t does not affect the values of n_t.

Model class:
For estimation purposes, we consider the following predictor corresponding to an Lth order FIR model

$$\hat{y}_t(\boldsymbol{\theta}) = G(z,\boldsymbol{\theta})u_t = \sum_{k=1}^{L} \theta_k z^{-k} u_t = \boldsymbol{\phi}_t^T \boldsymbol{\theta}, \tag{12}$$

where $\boldsymbol{\phi}_t = [u_{t-1}, \cdots, u_{t-L}]^T$ and $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_L]^T$.

Objective:
Design the input signal sequence u_t and provide a guaranteed confidence region for $\boldsymbol{\theta}^0 = [\theta_1^0, \cdots, \theta_L^0]^T$ using N input-output data $\{u_t, y_t\}_{t=1,\cdots,N}$.

B. Construction of confidence regions

First we design the input signal and determine confidence regions $\boldsymbol{\Theta}_s$ based on the sign of the correlation between the prediction error $\epsilon_t(\boldsymbol{\theta})$ and the input u_{t-s} for $s \in \{1, \ldots, L\}$.

Input design:

(**D1**) The input signal sequence $\{u_t\}$, for $t = 1, \cdots, N$, is independent and has equal probability 0.5 of being larger or smaller than zero.

Procedure for the construction of $\boldsymbol{\Theta}_s$:

(1) Compute the prediction errors

$$\epsilon_t(\boldsymbol{\theta}) = y_t - \hat{y}_t(\boldsymbol{\theta}) = y_t - \phi_t^T \boldsymbol{\theta} \tag{13}$$

for $t = 1 + L, 2 + L, \cdots, K + L = N$.

(2) Select an integer $s \in \{1, \cdots, L\}$ and compute

$$f_{t-s,s}(\boldsymbol{\theta}) = \text{sign}\big[u_{t-s}\epsilon_t(\boldsymbol{\theta})\big] \tag{14}$$

for $t = 1+L, \cdots, K+L$.

(3) Let $\mathbf{G}(K) = \{\mathbf{I}_i,\ i = 0, \cdots, M-1\}$ be a collection of subsets of $\{1, 2, \cdots, K\}$ forming a group with respect to the symmetric difference and let without loss of generality $\mathbf{I}_0 = \emptyset$. Compute the empirical correlations

$$\bar{g}_{i,s}(\boldsymbol{\theta}) = \sum_{t-L \in \mathbf{I}_i} f_{t-s,s}(\boldsymbol{\theta}),\ i = 0, \cdots, M-1. \tag{15}$$

(4) Add a small random number ν_i uniformly distributed on $[-a,\ a]$ with $a < 0.5$ to each correlation functions apart from the zero function $\bar{g}_{0,s}(\boldsymbol{\theta}) \equiv 0$

$$g_{i,s}(\boldsymbol{\theta}) = \bar{g}_{i,s}(\boldsymbol{\theta}) + \nu_i, \quad i = 1, \cdots, M-1. \tag{16}$$

The addition of ν_i prevents ties from occurring in the next step.

(5) Select an integer q in the interval $[1, (M+1)/2)$ and find the region $\boldsymbol{\Theta}_s$ such that at least q of the $g_{i,s}(\boldsymbol{\theta})$ functions are greater than the zero-function $g_{0,s}(\boldsymbol{\theta}) \equiv 0$ and at least q are smaller than $g_{0,s}(\boldsymbol{\theta}) \equiv 0$.

For $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, $\text{sign}\big[u_{t-s}\epsilon_t(\boldsymbol{\theta})\big]$ is a sequence of independent random variables with symmetric distribution around zero (see [7]). Therefore, it is unlikely that nearly all of the correlations functions $g_{i,s}(\boldsymbol{\theta})$ are positive or negative corresponding to the true value $\boldsymbol{\theta}^0$, and those regions in parameter space where this happens are therefore excluded from the confidence regions in point (5) of the procedure. The following theorem gives the exact probability that θ^0 belongs to the constructed region.

Theorem 1: In addition to $(\mathbf{A1})$*, assume that*

$$(\mathbf{A2}) \quad \Pr\big\{\epsilon_t(\boldsymbol{\theta}^0) = 0\big\} = 0.$$

Then,

$$\Pr\big\{\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}_s\big\} = 1 - 2 \cdot q/M. \tag{17}$$

Proof: See [7]. ■

Remark 1: The assumption $(\mathbf{A2})$ in Theorem 1 is mild. It is typically only violated when there is no undermodelling and n_t takes on the value 0 with non-zero probability. ■

The evaluation (17) is non-conservative in the sense that $1 - 2 \cdot q/M$ is the exact probability, not a lower bound. For each $s \in \{1, \ldots, L\}$, the set $\boldsymbol{\Theta}_s$ is a non-asymptotic confidence set for $\boldsymbol{\theta}^0$. However, each one of these sets can be unbounded in some directions of the parameter space, and therefore not particularly useful. A useful confidence set $\hat{\boldsymbol{\Theta}}$ is obtained by intersecting all the sets $\boldsymbol{\Theta}_s$ for $s = 1, \cdots, L$, i.e.

$$\hat{\boldsymbol{\Theta}} = \bigcap_{s=1}^{L} \boldsymbol{\Theta}_s. \tag{18}$$

From Theorem 1 it follows that

Theorem 2: Under the assumptions of Theorem 1,

$$\Pr\big\{\boldsymbol{\theta}^0 \in \hat{\boldsymbol{\Theta}}\big\} \geq 1 - 2 \cdot L \cdot q/M. \tag{19}$$

By adding some further statistical assumptions on the input and noise sequence, we can show that the constructed region concentrates around the true parameter $\boldsymbol{\theta}^0$ as the number of data points increases [7].

Remark 2 (Errors-in-variables model): The fact that Theorem 1 and Theorem 2 hold with a minor assumption $(\mathbf{A1})$ on n_t provides much flexibility in practical situations where the input to the system is also corrupted with an unknown noise sequence $\{\eta_t\}$ independent of $\{u_t\}$. Let

$$r_t = u_t + \eta_t \tag{20}$$

as shown in Fig. 4. Substituting (20) into (10) yields

$$y_t = G^0(z)r_t + n_t = G^0(z)u_t + e_t \tag{21}$$

where $e_t \triangleq G^0(z)\eta_t + n_t$.

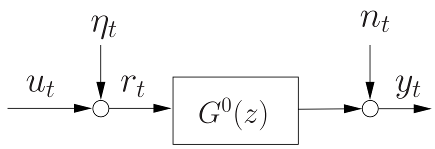

Fig. 4. Errors-in-variables model

By treating e_t as the new measurement noise sequence, we can construct confidence regions for the system parameters by following the same procedures as in Section II-B with the data set $\{u_t, y_t\}_{t=1,\dots,N}$. ■

III. GENERALIZATIONS

It is well known [3], [5] that using the pulse basis functions for identification of moderately damped systems or of systems with high sampling rates leads to approximations of high order. To deal with these situations, several orthonormal functions have been suggested which incorporate prior system information, e.g., the Laguerre functions [3] and the Kautz functions [4], both of which are special cases of the generalized orthonormal basis functions introduced in [5] and [6]. In this section, we first briefly describe these generalized orthonormal basis functions and then extend the results from the previous section to cover series expansion in these basis functions.

A. Generalized orthonormal basis functions

The theorem below from [5] describes the generalized orthonormal basis functions.

Theorem 3: Let $G_b(z)$ be a stable all-pass transfer function having an internally balanced realization[1] $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ *of order n_b. Denote*

$$\mathbf{V}_k(z) = \mathbf{V}_{k-1}(z)G_b(z) \quad \text{with } \mathbf{V}_0(z) = z(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}.$$

Then the sequence of scalar rational functions $\{\mathbf{e}_i^T \mathbf{V}_k(z)\}_{i=1,\cdots,n_b;k=0,\dots\infty}$ forms an orthonormal basis for the Hilbert space $\mathcal{H}_2$. Here $\mathbf{e}_i$ is i-th Euclidean basis vector in $\mathbb{R}^{n_b}$.

Note here that the basis functions constructed above are complete in $\mathcal{H}_2$ since the all-pass transfer function $G_b(z)$ is stable (see [8]).

One can construct an all-pass transfer function $G_b(z)$ from any given set of poles, and thus the resulting basis can incorporate dynamics of any complexity, combining, for example, both fast and slow dynamics in damped and resonant modes (see Proposition 7.1 of [5].) Corollary 1 below follows directly from Theorem 3.

Corollary 1: Let $G_b(z)$ be a stable all-pass transfer function of order n_b with a corresponding sequence of basis functions $\mathbf{V}_k(z)$ as in Theorem 3. Then for every strictly proper stable transfer function $H \in \mathcal{H}_2$ there exists a unique sequence $\mathbf{L} = \{\mathbf{L}_k\}_{k=0,1,\dots} \in \ell_2^{1\times n_b}[0,\infty)$, such that

$$H(z) = z^{-1}\sum_{k=0}^{\infty} \mathbf{L}_k \mathbf{V}_k(z). \tag{22}$$

We refer to $\mathbf{L}_k$ as the orthonormal expansion coefficients of $H(z)$.

The pulse, Laguerre, and Kautz functions are special cases of the generalized orthonormal basis functions as shown next.

Pulse Functions: Using the all-pass transfer function $G_b(z) = z^{-1}$ with minimal balanced realization $(0, 1, 1, 0)$, we obtain the standard pulse basis

$$V_k(z) = G_b^k(z) = z^{-k}.$$

Laguerre Functions: Using the all-pass transfer function $G_b(z) = (1 - az)/(z - a)$ for some real-valued a with $|a| < 1$, and balanced realization

$$(A, B, C, D) = (a, \sqrt{1-a^2}, \sqrt{1-a^2}, -a),$$

the Laguerre basis results [3]: $V_k(z) = \sqrt{1-a^2}\, z \frac{(1-az)^k}{(z-a)^{k+1}}$

Kautz Functions: Using the all-pass transfer function $G_b(z) = \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c}$ for some real-valued b, c with $|c|, |b| < 1$, and a balanced realization

$$\mathbf{A} = \begin{bmatrix} b & \sqrt{1-b^2} \\ c\sqrt{1-b^2} & -bc \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ \sqrt{1-c^2} \end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix} \gamma_2 & \gamma_1 \end{bmatrix}, \qquad D = -c$$

with $\gamma_1 = -b\sqrt{1-c^2}$ and $\gamma_2 = \sqrt{(1-c^2)(1-b^2)}$,

$$z^{-1}\mathbf{V}_0(z) = \frac{\sqrt{1-c^2}}{z^2 + b(c-1)z - c} \begin{bmatrix} \sqrt{1-b^2} \\ z - b \end{bmatrix},$$

we obtain the Kautz functions [4].

B. Generalized FIR models

In this section we convert models, which are series expansions in the above generalized basis functions, into FIR models using a filtering procedure.

Using (22) and denoting $\boldsymbol{\mathcal{B}}_{k+1}(z) \triangleq z^{-1}\mathbf{V}_k(z)$ and $\mathbf{L}_{k+1} \triangleq \boldsymbol{\theta}_k^0 \in \mathbb{R}^{1\times n_b}$, we can represent the true transfer function as

$$G^0(z) = \sum_{k=1}^{\infty} \boldsymbol{\theta}_k^0 \boldsymbol{\mathcal{B}}_k(z) \tag{23}$$

where the basis functions have the property

$$\boldsymbol{\mathcal{B}}_k(z) = \boldsymbol{\mathcal{B}}(z) \cdot \mathcal{A}^{k-1}(z) \tag{24}$$

[1] In [5], balanced realization was used simply as a technique to generate orthonormal basis functions from an all-pass transfer function $G_b(z)$.

with $\boldsymbol{\mathcal{B}}(z) \triangleq (z\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}, \quad \mathcal{A}(z) \triangleq G_b(z)$. Then, a linear time-invariant system can be represented as

$$y_t = G^0(z)u_t + n_t = \sum_{k=1}^{\infty} \boldsymbol{\theta}_k^0 \boldsymbol{\mathcal{B}}_k(z)u_t + n_t. \tag{25}$$

In order to use the procedure for constructing confidence regions developed in Section II, we create a new input-output data set $\{\tilde{\mathbf{u}}_t, \tilde{y}_t\}_{t=1,\cdots,N}$ from the original data set $\{u_t, y_t\}_{t=1,\cdots,N}$ such that the new data set satisfies

$$\tilde{y}_t = \sum_{k=1}^{\infty} \boldsymbol{\theta}_k^0 z^{-k}\tilde{\mathbf{u}}_t + \tilde{n}_t. \tag{26}$$

The filtered data set $\{\tilde{\mathbf{u}}_t, \tilde{y}_t\}$ is constructed starting from the last elements $\{\tilde{\mathbf{u}}_N, \tilde{y}_N\}$ (at time $t = N$) to the first elements $\{\tilde{\mathbf{u}}_1, \tilde{y}_1\}$ through successive filtering with the all-pass filter $\mathcal{A}(z)$ as described below:

For i = N: Take

$$\tilde{\mathbf{u}}_N \triangleq \{u_t\}|_{t=N} \times \mathbf{1}, \quad \tilde{y}_N \triangleq \{y_t\}|_{t=N}.$$

Here $\mathbf{1}$ is a vector of all ones in $\mathbb{R}^{n_b}$.

For i = N-1,$\cdots$,1: Apply filtering N-i times and collect the last elements

$$\tilde{\mathbf{u}}_i \triangleq \left[\mathcal{A}^{N-(i+1)}(z)\boldsymbol{\mathcal{B}}(z)\right]\{u_t\}\big|_{t=N}, \tag{27}$$

$$\tilde{y}_i \triangleq \left[\mathcal{A}^{N-i}(z)\{y_t\}\right]\big|_{t=N}. \tag{28}$$

By applying the filtering above, we obtain

$$\begin{aligned}
\tilde{y}_N &= \{y_t\}|_{t=N} \\
&= \boldsymbol{\theta}_1^0\boldsymbol{\mathcal{B}}_1\{u_t\}|_{t=N} + \boldsymbol{\theta}_2^0\boldsymbol{\mathcal{B}}_2\{u_t\}|_{t=N} + \cdots + \{n_t\}|_{t=N} \\
&= \boldsymbol{\theta}_1^0\boldsymbol{\mathcal{B}}\{u_t\}|_{t=N} + \boldsymbol{\theta}_2^0\mathcal{A}\boldsymbol{\mathcal{B}}\{u_t\}|_{t=N} + \cdots + n_N \\
&= \boldsymbol{\theta}_1^0\tilde{\mathbf{u}}_{N-1} + \boldsymbol{\theta}_2^0\tilde{\mathbf{u}}_{N-2} + \cdots + \tilde{n}_N
\end{aligned}$$

and similarly, for $i = N-1, \cdots, 1$,

$$\begin{aligned}
\tilde{y}_i &= \mathcal{A}^{N-i}\{y_t\}|_{t=N} \\
&= \boldsymbol{\theta}_1^0\mathcal{A}^{N-i}\boldsymbol{\mathcal{B}}_1\{u_t\}|_{t=N} + \boldsymbol{\theta}_2^0\mathcal{A}^{N-i}\boldsymbol{\mathcal{B}}_2\{u_t\}|_{t=N} + \cdots \\
&\quad + \mathcal{A}^{N-i}\{n_t\}|_{t=N} \\
&= \boldsymbol{\theta}_1^0\tilde{\mathbf{u}}_{i-1} + \boldsymbol{\theta}_2^0\tilde{\mathbf{u}}_{i-2} + \cdots + \tilde{n}_i.
\end{aligned}$$

Therefore, we can represent the system in terms of the pulse basis functions

$$\tilde{y}_t = \sum_{k=1}^{\infty} \boldsymbol{\theta}_k^0 z^{-k}\tilde{\mathbf{u}}_t + \tilde{n}_t \tag{29}$$

with $\tilde{n}_i \triangleq \mathcal{A}^{N-i}(z)\{n_t\}|_{t=N}$. The corresponding predictor and prediction error become

$$\hat{\tilde{y}}_t(\boldsymbol{\theta}) = \sum_{k=1}^{L} \boldsymbol{\theta}_k z^{-k}\tilde{\mathbf{u}}_t, \;\; \tilde{\epsilon}_t(\boldsymbol{\theta}) = \tilde{y}_t - \hat{\tilde{y}}_t(\boldsymbol{\theta}). \tag{30}$$

Now we design the following new input signal u_t and make a strengthening assumption on the noise n_t:

(**D2**) The input $\{u_t\}$ is a white gaussian sequence with spectral density Φ_u.

(**A3**) The the noise $\{n_t\}$ is a white gaussian sequence independent of the input u_t.

Then,

$$\begin{aligned}
E\{\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_j^T\} &= [\; N-i-1 \triangleq k, \;\; N-j-1 \triangleq l \;] \\
&= E\{\boldsymbol{\mathcal{B}}_k(z)\{u_t\}|_{t=N} \cdot \boldsymbol{\mathcal{B}}_l^T(z)\{u_t\}|_{t=N}\} \\
&= \frac{\Phi_u}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{\mathcal{B}}_k(e^{j\omega})\boldsymbol{\mathcal{B}}_l^T(e^{-j\omega})d\omega = \Phi_u\mathbf{I}\delta(i-j)
\end{aligned} \tag{31}$$

where we have used the Parseval's relationship and the orthonormality. Hence, the filtered input sequence $\{\tilde{\mathbf{u}}_t\}$ is uncorrelated and also independent, since $\{\tilde{\mathbf{u}}_t\}$ is gaussian. Furthermore, $\{\tilde{\mathbf{u}}_t\}$ and $\{\tilde{n}_t\}$ are strict sense stationary and ergodic (see [7]). Therefore, from the observation that the new input $\{\tilde{\mathbf{u}}_t\}$ and the new noise sequence $\{\tilde{n}_t\}$ satisfy assumptions (**A1**) - (**A2**) in Section II, the procedure for construction of confidence regions for $\boldsymbol{\theta}_k^0, \; k = 1, \cdots, L$ developed in Section II-B can be applied to (29) with the new data set $\{\tilde{\mathbf{u}}_t, \tilde{y}_t\}$ (as summarized in Corollary 2 below), and the convergence property explained in Section II-B also holds for the system (29)(see [7]).

Corollary 2: Under the input design $(\mathbf{D2})$ *and the assumption* $(\mathbf{A3})$*, the set* $\boldsymbol{\Theta}_s$ *obtained by applying the procedure in Section II-B to the system* (29) *with the filtered data* (27)- (28) *has the property that*

$$\Pr\{\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}_s\} = 1 - 2 \cdot q/M,$$

and the set $\hat{\boldsymbol{\Theta}} = \cap_{s=1}^{L}\boldsymbol{\Theta}_s$ *has the property that*

$$\Pr\{\boldsymbol{\theta}^0 \in \hat{\boldsymbol{\Theta}}\} \geq 1 - 2 \cdot L \cdot q/M.$$

Remark 3 (Effects of initial conditions): Performing the successive filtering described above requires information about past input and output $\{u_t, y_t\}_{t\leq 0}$. We can decompose the input and the output as $y_t = y_t^+ + y_t^-, \; u_t = u_t^+ + u_t^-$ where $(\cdot)_t^+ \triangleq (\cdot)_t$ for $t > 0$ and $(\cdot)_t^+ = 0$ for $t \leq 0$, and $(\cdot)_t^- \triangleq (\cdot)_t$ for $t \leq 0$ and $(\cdot)_t^- = 0$ for $t > 0$. Then, the system equation (25) can be written as

$$\begin{aligned}
y_t^+ &= G^0(z)u_t^+ + \left[n_t + G^0(z)u_t^- - y_t^-\right] \\
&= G^0(z)u_t^+ + \bar{n}_t
\end{aligned}$$

with $\bar{n}_t \triangleq n_t + G^0(z)u_t^- - y_t^-$. By treating $\bar{n}_t$ as a new noise sequence, we can construct confidence regions for the parameters of the system by following the previous procedure using the data set $\{u_t^+, y_t^+\}$. Generally, the new input sequence obtained by repeated filtering of $\{u_t^+\}$ is not an independent sequence due to the zero initial condition. The exception is the pulse basis functions for which the sequence will be independent. However, for a filter $\boldsymbol{\mathcal{B}}_k(z)$ with a fast-decaying impulse response, the magnitude of the tail is so small that the new input $\{\tilde{\mathbf{u}}_k^+\}$ in practice can be treated as an independent sequence. ■

IV. Numerical Example

In order to numerically demonstrate the effectiveness of the algorithms described in previous sections, we consider the following discrete-time system

$$y_t = G^0(z)u_t + n_t. \tag{32}$$

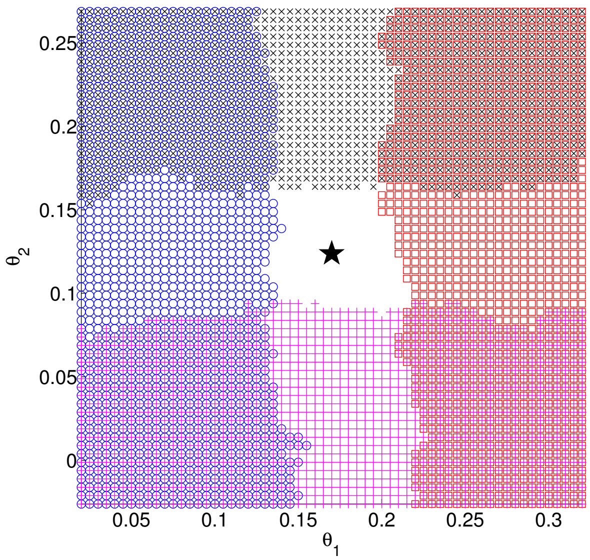

Fig. 5. Non-asymptotic 95% confidence region for (θ_1^0, θ_2^0) (blank region) using 511 filtered data points. ★ = true parameter, Laguerre case

The transfer function $G^0(z)$ is given by

$$G^0(z) = \frac{0.03555z + 0.02465}{(z - 0.90483)(z - 0.36787)} \quad (33)$$

which was obtained from a continuous-time system

$$\frac{1}{(10s + 1)(s + 1)} \quad (34)$$

by discretizing with zero-order-hold and sampling period 1 second. The input sequence $\{u_t\}$ and the noise sequence $\{n_t\}$ were zero-mean white Gaussian sequences with variance 1 and 0.05, respectively. 5000 data points were collected.

The coefficients of the pulse basis functions slowly decay, but for the Laguerre basis function the two terms are sufficient to get a good approximation. Therefore, we use the 2nd order Laguerre model for prediction

$$\hat{y}_t = \theta_1 \mathcal{B}_1(z) u_t + \theta_2 \mathcal{B}_2(z) u_t \quad (35)$$

with $a = 0.8$ the pole location of the Laguerre basis functions between the true system poles.

To obtain confidence regions for θ_1^0 and θ_2^0, the successive filtering explained in Section III-B was applied to the original data set $\{u_t, y_t\}_{t=1,\cdots,N}$ generated from (32) and a filtered data set $\{\tilde{u}_t, \tilde{y}_t\}_{t=1,\cdots,N}$ was obtained. The initial conditions were set to zero as in Remark 3.

The last 511 data points out of the 5000 filtered data points were used. We computed

$$f_{t-1,1}(\boldsymbol{\theta}) = \text{sign}\big[\tilde{u}_{t-1}\epsilon_t(\boldsymbol{\theta})\big],$$
$$f_{t-2,2}(\boldsymbol{\theta}) = \text{sign}\big[\tilde{u}_{t-2}\epsilon_t(\boldsymbol{\theta})\big],$$

for $t = 4490, \cdots, 5000$, and

$$g_{i,1}(\boldsymbol{\theta}) = \sum_{t-2 \in \mathbf{I}_i} f_{t-1,1}(\boldsymbol{\theta}) + \nu_{i,1} \quad i = 0, \cdots, M-1$$
$$g_{i,2}(\boldsymbol{\theta}) = \sum_{t-2 \in \mathbf{I}_i} f_{t-1,2}(\boldsymbol{\theta}) + \nu_{i,2} \quad i = 0, \cdots, M-1$$

where $\nu_{i,1}$ and $\nu_{i,2}$ were uniformly distributed on $[-0.1,\ 0.1]$. We excluded the regions in parameter space where at most five (out of the $M = 512$) $g_{i,1}$ and $g_{i,2}$ functions were positive or negative. The obtained confidence region is the blank area in Fig. 5. The region constructed this way has a probability of at least $1 - 2 \cdot 2 \cdot 6/512 = 0.9531$ of containing the true parameter. The true value is marked with ★. The regions where at most five $g_{i,1}(\boldsymbol{\theta})$ functions were negative are marked with ○, and the regions where at most five were positive are marked with □. Likewise for $g_{i,2}(\boldsymbol{\theta})$, where $+$ and $\times$ represent the regions where at most five values of $g_{i,2}(\boldsymbol{\theta})$ were negative or positive. As we can see, each step in the construction of the confidence region excludes a particular region.

V. Concluding Remarks

In this paper, we have extended the LSCR algorithm developed in [1], [2] for constructing non-asymptotic confidence regions to the case where undermodelling is present. The systems are approximated by generalized orthonormal basis functions models, and by applying the sign-function in the computations of the correlation functions, guaranteed non-asymptotic confidence regions can be constructed. The method was first developed for FIR models and then extended to models represented by generalized orthonormal basis functions through a filtering procedure.

Acknowledgments

This research of M. C. Campi was supported by MIUR under the project "Identification and Adaptive Control for Industrial Systems." The research of S. Ko and E. Weyer was support by the Australian Research Council under the Discovery Grant Scheme, Project DP0558579.

References

[1] M. C. Campi and E. Weyer, "Guaranteed non-asymptotic confidence regions in system identification," *Automatica*, vol. 41, no. 10, pp. 1751–1764, Oct. 2005.

[2] ——, "Non-asymptotic confidence sets for input-output transfer functions," in *Proc. 45th IEEE Conference on Decision and Control*, San Diego, USA, 2006, pp. 157–162.

[3] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. Automat. Contr.*, vol. 36, no. 5, pp. 551–562, May 1991.

[4] ——, "System identification using Kautz models," *IEEE Trans. Automat. Contr.*, vol. 39, no. 6, pp. 1276–1282, Jun. 1994.

[5] P. S. C. Heuberger, P. M. J. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 40, no. 3, pp. 451–465, Mar. 1995.

[6] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor, "System identification with generalized orthonormal basis functions," *Automatica*, vol. 31, no. 12, pp. 1821–1834, Dec. 1995.

[7] M. C. Campi, S. Ko, and E. Weyer, "Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics," Department of Electrical and Electronic Engineering, The University of Melbourne, Tech. Rep., 2007.

[8] P. S. Heuberger, T. J. De Hoog, P. M. J. Van den hof, and B. Wahlberg, "Orthonormal basis functions in time and frequency domain: HAMBO transform theory," *SIAM J. Control Optim.*, vol. 42, no. 4, pp. 1347–1373, 2003.