Proceedings of the
15th IFAC Symposium on System Identification
Saint-Malo, France, July 6-8, 2009

Online Algorithms for the Construction of Guaranteed Confidence Sets for the Parameters of Time-varying Systems ⋆

Kelvin J. Layton * **Erik Weyer** * **Marco C. Campi** **

* *Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. (email: {laytonk,e.weyer}@ee.unimelb.edu.au)*
** *Department of Electrical Engineering and Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy (e-mail: campi@ing.unibs.it)*

Abstract: The "Leave-out Sign-dominant Correlation Regions" (LSCR) algorithm is extended to deliver a guaranteed confidence set for the parameters of a time-varying system at any given time. The algorithm is derived by assuming that an upper bound on the parameter variation is available, and the delivered confidence set is valid without any prior knowledge of the noise. Simulation examples are provided to illustrate the performance of the algorithm.

Keywords: Time-varying systems; System identification; Parameter estimation; Recursive algorithms; Uncertainty

1. INTRODUCTION

System identification deals with the problem of building mathematical models of dynamical systems from observed data (Ljung [1999], Söderström and Stoica [1988]). It is 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 provides an 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 derive an algorithm for tracking of parameters of a time varying system. Under the assumption that an upper bound on the variability of the system parameters is known, the algorithm returns a confidence region that tracks the true parameter variation through time. At each instant of time the confidence region is guaranteed to contain the true parameter with a user chosen probability. This holds true irrespective of the noise level corrupting the observations, and no prior knowledge about the noise process is required for implementation of the algorithm.

The algorithm constructed in this paper builds on previous work on time-invariant system identification (Campi and Weyer [2005, 2006a,b]), along the LSCR (Leave-out Sign-dominant Correlation Regions) approach.

The paper is organised as follows. In the next section we give a brief overview of the LSCR algorithm for time-invariant systems, before we extend it to time-varying systems in Section 3. Simulation results demonstrating the method are presented in Section 4.

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

2. LSCR FOR TIME-INVARIANT SYSTEMS

In this section we briefly review the LSCR algorithm for time-invariant systems.

2.1 Data generating system and predictors

The data generating system is given by

$$y_t = \boldsymbol{\phi}_t^T \boldsymbol{\theta}^0 + w_t \tag{1}$$

where

$$\boldsymbol{\phi}_t = [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-1}, \ldots, u_{t-n_b}]^T$$
$$\boldsymbol{\theta}^0 = [a_1^0, \ldots, a_{n_a}^0, b_1^0, \ldots, b_{n_b}^0]^T$$

Here y_t and u_t are the output and input respectively and w_t is the noise affecting the system.

Assumptions

(1) The user can choose the input signal u_t, and the choice of u_t does not affect w_t. That is u_t and w_t are independent.
(2) The model orders n_a and n_b are known.

There are no assumptions on w_t. No upper bound on its magnitude is assumed, and it is allowed to have non-zero mean and any autocorrelation properties.

As a predictor we use

$$\hat{y}_t(\boldsymbol{\theta}) = \boldsymbol{\phi}_t^T \boldsymbol{\theta}$$

where

$$\boldsymbol{\theta} = [a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}]^T$$

and the prediction error is given by $\epsilon_t(\boldsymbol{\theta}) = y_t - \hat{y}_t(\boldsymbol{\theta})$.

2.2 Main idea

The LSCR algorithm introduced by Campi and Weyer [2005] generates a region in parameter space that has a guaranteed probability of containing the true parameter $\boldsymbol{\theta}^0$ using a finite number of data points N. The main idea is to compute empirical estimates of the correlation between the prediction error $\epsilon_t(\boldsymbol{\theta})$ and u_{t-k} using different subsets of the data. If u_t is chosen as an independent sequence symmetrically distributed around zero, one observes that for $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, $\epsilon_t(\boldsymbol{\theta})u_{t-k} = w_t u_{t-k}$ is also an independent sequence symmetrically distributed around zero. Hence, for the true value $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ it is very unlikely that nearly all of the estimates of the correlation will be positive or that nearly all of them will be negative. The values of $\boldsymbol{\theta}$ for which this occurs are therefore excluded from the confidence set, hence the name Leave-out Sign-dominant Correlation Regions.

2.3 The LSCR algorithm

Input design. Let u_t, $t = 1-n, \ldots, N$, be an iid sequence of random variables symmetrically distributed around 0, and let $n = n_a + n_b$ be the number of parameters in the parameter vector $\boldsymbol{\theta}$.

LSCR Algorithm.

(S1′) Compute the prediction error

$$\varepsilon_t(\boldsymbol{\theta}) = y_t - \hat{y}_t(\boldsymbol{\theta}) = y_t - \boldsymbol{\phi}_t^T\boldsymbol{\theta} \qquad t = 1, \ldots, N$$

(S2′) Form the vector

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

of length $n = n_a + n_b$ and compute the vector

$$\boldsymbol{f}_t(\boldsymbol{\theta}) = \boldsymbol{\xi}_t\epsilon_t(\boldsymbol{\theta}), \quad t = 1, \ldots, N.$$

(S3′) Select an integer M and construct M binary ($\{0,1\}$-valued) stochastic strings of length N as follows: Let $h_{0,1}, \ldots, h_{0,N} = 0, \ldots, 0$ be the string of all zeros. Every element of the remaining strings takes the values 0 or 1 each with probability 0.5, and the elements are independent of each other. Moreover, each string is constructed independently of previous strings. Name the constructed non-zero strings $h_{1,1}, \ldots, h_{1,N}$; $h_{2,1}, \ldots, h_{2,N}$;…; $h_{M-1,1}, \ldots, h_{M-1,N}$. Compute

$$\begin{aligned}\boldsymbol{g}_i(\boldsymbol{\theta}) &= \sum_{t=1}^{N} h_{i,t} \cdot \boldsymbol{f}_t(\boldsymbol{\theta}) \\ &= \sum_{t=1}^{N} h_{i,t} \cdot \boldsymbol{\xi}_t\epsilon_t(\boldsymbol{\theta}), \quad i = 0, \ldots, M-1.\end{aligned}$$

Note that $\boldsymbol{g}_0(\boldsymbol{\theta}) \equiv \mathbf{0}$.

(S4′) Let $g_i^k(\boldsymbol{\theta})$ denote the kth element of the vector $\boldsymbol{g}_i(\boldsymbol{\theta})$, $k = 1, \ldots, n$. Select an integer q in the interval $[1, M/2n]$. Construct the regions $\hat{\boldsymbol{\Theta}}_N^k$ such that at least q of the $g_i^k(\boldsymbol{\theta})$ functions are strictly larger than $g_0^k(\boldsymbol{\theta}) \equiv 0$ and at least q are strictly smaller than $g_0^k(\boldsymbol{\theta}) \equiv 0$. The confidence set is given by

$$\hat{\boldsymbol{\Theta}}_N = \bigcap_{k=1}^{n} \hat{\boldsymbol{\Theta}}_N^k. \tag{2}$$

Point (S4′) of the algorithm implements the main idea described in Section 2.2. That is, for $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ the functions $g_i^k(\boldsymbol{\theta}) = \sum_{t=1}^{N} h_{i,t} \cdot u_{t-k}\epsilon_t(\boldsymbol{\theta})$ take on positive and negative values at random since $g_i^k(\boldsymbol{\theta}^0) = \sum_{t=1}^{N} h_{i,t} \cdot u_{t-k}w_t$. It is therefore unlikely that only a small fraction of them are positive or negative, and the algorithm excludes the regions in parameter space where this happens. The probability that $\boldsymbol{\theta}^0$ belongs to each of the $\hat{\boldsymbol{\Theta}}_N^k$ is given in the next theorem.

Theorem 1. Consider a $k \in \{1, \ldots, n\}$ and assume that $\Pr\{g_i^k(\boldsymbol{\theta}^0) = 0\} = 0$, $i \neq 0$. Then,

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

where $\hat{\boldsymbol{\Theta}}_N^k$ is constructed in point (S4′) of the algorithm above, and q and M are introduced in points (S4′) and (S3′) respectively.

Since $\Pr\{\boldsymbol{\theta}^0 \notin \hat{\boldsymbol{\Theta}}_N^k\} = 2q/M$ for every $k \in \{1, \ldots, n\}$, it follows that $\Pr\{\boldsymbol{\theta}^0 \notin \bigcup_{k=1}^{n} \hat{\boldsymbol{\Theta}}_N^k\} \leq 2nq/M$. Note that the latter is an inequality since the events $\{\boldsymbol{\theta}^0 \notin \hat{\boldsymbol{\Theta}}_N^k\}, k = 1, \ldots, n$ may overlap. It follows that

Corollary 2. Under the assumptions in Theorem 1,

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

where $\hat{\boldsymbol{\Theta}}_N$ is given by (2).

Theorem 1 and Corollary 2 are valid regardless of the noise characteristics. The size of the confidence region will, however, depend on the noise level, such that the regions become larger when the noise level increases. The LSCR algorithm automatically adjusts for an increased noise level in (S4′) by keeping more values of $\boldsymbol{\theta}$ in the confidence set.

3. LSCR FOR TIME-VARYING SYSTEM

In this section, we extend the LSCR algorithm described above to provide confidence sets for the parameters of a time-varying system.

3.1 Data generating system

The data generating system is now given by

$$y_t = \boldsymbol{\phi}_t^T\boldsymbol{\theta}_t^0 + w_t \tag{3}$$

where

$$\begin{aligned}\boldsymbol{\phi}_t &= [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-1}, \ldots, u_{t-n_b}]^T \\ \boldsymbol{\theta}_t^0 &= [a_{1,t}^0, \ldots, a_{n_a,t}^0, b_{1,t}^0, \ldots, b_{n_b,t}^0]^T\end{aligned}$$

This is the same as (1) except that the true parameter $\boldsymbol{\theta}_t^0$ varies with time.

The assumptions on the noise, w_t, and input, u_t, are the same as in the time-invariant case. In particular, we assume no knowledge about the noise.

We assume that the variation in $\boldsymbol{\theta}_t^0$ is bounded by $|\boldsymbol{\theta}_{t+1}^0 - \boldsymbol{\theta}_t^0| \leq \boldsymbol{K}$ where $\boldsymbol{K} = \left[\Delta a_1^0, \ldots, \Delta a_{n_a}^0, \Delta b_1^0, \ldots, \Delta b_{n_b}^0\right]^T$, $|\cdot|$ denotes element-wise absolute value and $\leq$ represents an element-wise comparison. For example, Δa_1^0 is the magnitude of the maximum change of the true a_1^0 parameter

between time steps. With this assumption, we aim to find a region $\hat{\boldsymbol{\Theta}}_t$, that contains the true parameter of the time-varying system at time t with a given user chosen probability.

3.2 Generalisations for time-varying systems

The algorithm for time-invariant systems works because for the true parameter, $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, $\varepsilon_t(\boldsymbol{\theta}^0)u_{t-k} = w_t u_{t-k}$. Thus, $g_i^k(\boldsymbol{\theta}^0) = \sum_{t=1}^N h_{i,t} w_t u_{t-k}$. These functions have the desired ordering property exploited in step (S4′) because w_t and u_{t-k} are independent and u_{t-k} is iid and symmetrically distributed around 0. For time-varying systems, $\boldsymbol{\theta}_N^0$ is the true parameter at time N and $\varepsilon_t(\boldsymbol{\theta}_N^0)u_{t-k} \neq w_t u_{t-k}$ for $t \neq N$. Instead we have $\varepsilon_t(\boldsymbol{\theta}_t^0)u_{t-k} = w_t u_{t-k}$ and the functions

$$\sum_{t=1}^{N} h_{i,t}\varepsilon_t(\boldsymbol{\theta}_t^0)u_{t-k} \tag{4}$$

have the desired ordering property. As the goal is to find a confidence set for $\boldsymbol{\theta}_N^0$ at time N we rewrite $\sum_{t=1}^N h_{i,t}\varepsilon_t(\boldsymbol{\theta}_t^0)u_{t-k}$ in terms of $\sum_{t=1}^N h_{i,t}\varepsilon_t(\boldsymbol{\theta}_N^0)u_{t-k}$ and an extra term $C_{i,N}^k$ that can be bounded based on the maximum parameter change $\boldsymbol{K}$.

Additionally, due to the time-varying nature of the system, data from the distance past should be given less weight so the algorithm can track parameter changes. The standard approach, as in Ljung and Söderström [1983], is to introduce a forgetting factor, $\lambda < 1$, such that the above sum in (4) becomes

$$f_{i,N}^k = \sum_{t=1}^{N} \lambda^{N-t} h_{i,t}\varepsilon_t(\boldsymbol{\theta}_t^0)u_{t-k} = \sum_{t=1}^{N} \lambda^{N-t} h_{i,t} w_t u_{t-k} \tag{5}$$

which still possesses the desired ordering property.

Next we express $\varepsilon_t(\boldsymbol{\theta}_t^0)$ in terms of $\varepsilon_t(\boldsymbol{\theta}_N^0)$. The predictor and prediction error are

$$\begin{aligned}\hat{y}_t(\boldsymbol{\theta}) &= \boldsymbol{\phi}_t^T\boldsymbol{\theta}\\ \varepsilon_t(\boldsymbol{\theta}) &= y_t - \hat{y}_t(\boldsymbol{\theta}) = y(t) - \boldsymbol{\phi}_t^T\boldsymbol{\theta}\end{aligned}$$

The prediction error $\varepsilon_t(\boldsymbol{\theta}_t^0) = w_t$ can be rewritten as follows

$$\begin{aligned}\varepsilon_t(\boldsymbol{\theta}_t^0) &= \varepsilon_t(\boldsymbol{\theta}_N^0) + \varepsilon_t(\boldsymbol{\theta}_t^0) - \varepsilon_t(\boldsymbol{\theta}_N^0)\\ &= \varepsilon_t(\boldsymbol{\theta}_N^0) + \boldsymbol{\phi}_t^T(\boldsymbol{\theta}_N^0 - \boldsymbol{\theta}_t^0)\\ &= \varepsilon_t(\boldsymbol{\theta}_N^0) + \boldsymbol{\phi}_t^T\left(\sum_{j=t}^{N-1}(\boldsymbol{\theta}_{j+1}^0 - \boldsymbol{\theta}_j^0)\right)\end{aligned}$$

and multiplication with the input u_{t-k} gives,

$$w_t u_{t-k} = \left(\varepsilon_t(\boldsymbol{\theta}_N^0) + \boldsymbol{\phi}_t^T\left(\sum_{j=t}^{N-1}(\boldsymbol{\theta}_{j+1}^0 - \boldsymbol{\theta}_j^0)\right)\right)u_{t-k}$$

We now use this to expand (5)

$$\begin{aligned}f_{i,N}^k =& \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\varepsilon_t(\boldsymbol{\theta}_t^0)u_{t-k}\\ =& \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\varepsilon_t(\boldsymbol{\theta}_N^0)u_{t-k} +\\ & \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t^T\left(\sum_{j=t}^{N-1}(\boldsymbol{\theta}_{j+1}^0 - \boldsymbol{\theta}_j^0)\right)u_{t-k}\\ =& g_{i,N}^k(\boldsymbol{\theta}_N^0) + C_{i,N}^k\end{aligned} \tag{6}$$

where

$$g_{i,N}^k(\boldsymbol{\theta}) = \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\varepsilon_t(\boldsymbol{\theta})u_{t-k} \tag{7}$$

and

$$C_{i,N}^k = \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t^T\left(\sum_{j=t}^{N-1}(\boldsymbol{\theta}_{j+1}^0 - \boldsymbol{\theta}_j^0)\right)u_{t-k} \tag{8}$$

Note that $g_{i,N}^k(\boldsymbol{\theta})$ can be evaluated for an arbitrary $\boldsymbol{\theta}$, and that $C_{i,N}^k$ can be bounded due to the assumption $|\boldsymbol{\theta}_{j+1}^0 - \boldsymbol{\theta}_j^0| \leq \boldsymbol{K}$.

The goal is to find recursive algorithms for computing $g_{i,N}^k(\boldsymbol{\theta})$ and $\bar{C}_{i,N}^k$, where $\bar{C}_{i,N}^k$ is a bound such that $\bar{C}_{i,N}^k \geq |C_{i,N}^k|$. This means that $g_{i,N+1}^k(\boldsymbol{\theta})$ and $\bar{C}_{i,N+1}^k$ should be computed from $g_{i,N}^k(\boldsymbol{\theta})$ and $\bar{C}_{i,N}^k$ respectively, using a fixed finite memory and a fixed finite number of computations per time step.

3.3 Recursive algorithm for time-varying systems

Below we give the recursive LSCR algorithm. The detailed derivation is given in Appendix A. In the algorithm,

$$\bar{\boldsymbol{C}}_{i,N} = [\bar{C}_{i,N}^1, \ldots, \bar{C}_{i,N}^n]^T$$

and $\boldsymbol{D}_{i,N}$ is an $n \times n$ matrix used in the calculation of $\bar{\boldsymbol{C}}_{i,N}$. (S0) is an initialisation step, and the the following steps (S1) to (S4) are repeated for every 'new' data point (which we denote by $t = N + 1$).

LSCR algorithm for time-varying systems

(S0) Let

$$\begin{aligned}\boldsymbol{g}_{i,0}(\boldsymbol{\theta}) &= \boldsymbol{0} \qquad & i = 1, \ldots, M-1\\ \boldsymbol{C}_{i,0}(\boldsymbol{\theta}) &= \boldsymbol{0} \qquad & i = 1, \ldots, M-1\\ \boldsymbol{D}_{i,0}(\boldsymbol{\theta}) &= \boldsymbol{0} \qquad & i = 1, \ldots, M-1\end{aligned}$$

(S1) Compute the predictor and prediction error

$$\begin{aligned}\hat{y}_{N+1}(\boldsymbol{\theta}) &= \boldsymbol{\phi}_{N+1}^T\boldsymbol{\theta}\\ \varepsilon_{N+1}(\boldsymbol{\theta}) &= y_{N+1} - \hat{y}_{N+1}(\boldsymbol{\theta})\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\phi}_{N+1} &= [-y_N, \ldots, -y_{N+1-n_a}, u_N, \ldots, u_{N+1-n_b}]^T\\ \boldsymbol{\theta} &= [a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b}]^T\end{aligned}$$

(S2) Form the vector

$$\boldsymbol{\xi}_{N+1} = [u_N, \ldots, u_{N+1-n}]^T$$

(S3) Draw $h_{i,N+1} = 0$ or 1 with probability 0.5 each, for $i = 1, \ldots, M-1$, and calculate the vector functions $\boldsymbol{g}_{i,N+1}(\boldsymbol{\theta})$ and $\boldsymbol{C}_{i,N+1}$ for $i = 1, \ldots, M-1$ as follows:

$$\boldsymbol{g}_{i,N+1}(\boldsymbol{\theta}) = \lambda \boldsymbol{g}_{i,N}(\boldsymbol{\theta}) + h_{i,N+1}\varepsilon_{N+1}(\boldsymbol{\theta})\boldsymbol{\xi}_{N+1}$$
$$\boldsymbol{D}_{i,N+1} = \lambda \boldsymbol{D}_{i,N} + \lambda h_{i,N}\boldsymbol{\phi}_N \boldsymbol{\xi}_{N+1}^T$$
$$\bar{\boldsymbol{C}}_{i,N+1} = \lambda \bar{\boldsymbol{C}}_{i,N} + |\boldsymbol{D}_{i,N+1}|^T \boldsymbol{K}$$

where $|\cdot|$ denotes element-wise absolute value. Note that $\boldsymbol{g}_{0,N+1}(\boldsymbol{\theta}) \equiv \boldsymbol{0}$

(S4) Find the region $\hat{\boldsymbol{\Theta}}^k_{N+1}$ such that for all $\boldsymbol{\theta} \in \hat{\boldsymbol{\Theta}}^k_{N+1}$ at least q of the empirical correlation estimates $g^k_{i,N+1}(\boldsymbol{\theta})$ satisfy $g^k_{i,N+1}(\boldsymbol{\theta}) - \bar{C}^k_{i,N+1} < g^k_{0,N+1}(\boldsymbol{\theta})$ and at least q satisfy $g^k_{i,N+1}(\boldsymbol{\theta}) + \bar{C}^k_{i,N+1} > g^k_{0,N+1}(\boldsymbol{\theta})$, where $g^k_{i,N+1}(\boldsymbol{\theta})$ and $\bar{C}^k_{i,N+1}$ are the kth elements of $\boldsymbol{g}_{i,N+1}(\boldsymbol{\theta})$ and $\bar{\boldsymbol{C}}_{i,N+1}$ respectively. The confidence set for $t = N+1$ is given by

$$\hat{\boldsymbol{\Theta}}_{N+1} = \bigcap_{k=1}^{n} \hat{\boldsymbol{\Theta}}^k_{N+1}$$

3.4 Properties of the algorithm

The functions, $f^k_{i,N+1}$, from (6) have the required ordering property exploited in the LSCR algorithm. For the true parameter value $\boldsymbol{\theta}^0_{N+1}$ at time $N+1$ we have that

$$g^k_{i,N+1}(\boldsymbol{\theta}^0_{N+1}) - \bar{C}^k_{i,N+1} \le f^k_{i,N+1} \le g^k_{i,N+1}(\boldsymbol{\theta}^0_{N+1}) + \bar{C}^k_{i,N+1}$$

A confidence set is then constructed in the same manner as for time-invariant LSCR by excluding those regions in parameter space where $g^k_{i,N+1}(\boldsymbol{\theta}^0_{N+1}) - \bar{C}^k_{i,N+1}$ takes on positive values too many times and where $g^k_{i,N+1}(\boldsymbol{\theta}^0_{N+1}) + \bar{C}^k_{i,N+1}$ takes on negative values too many times.

Thus, the region constructed above will have at least the same guaranteed probability of containing the true parameter at any time t as stated in Theorem 1 and Corollary 2. These results are now restated in the context of the recursive algorithm for time-varying systems.

Theorem 3. Under the same assumptions as in Theorem 1, for any given time t, the confidence regions constructed in (S4) of the recursive LSCR algorithm in Section 3.3 have the properties that

$$\Pr\{\boldsymbol{\theta}^0_t \in \hat{\boldsymbol{\Theta}}^k_t\} \ge 1 - 2q/M$$

and

$$\Pr\{\boldsymbol{\theta}^0_t \in \hat{\boldsymbol{\Theta}}_t\} \ge 1 - 2nq/M$$

4. SIMULATION EXAMPLE

The following first-order system was simulated

$$y_t = -a^0_t y_{t-1} + b^0_t u_{t-1} + w_t \tag{9}$$

For this example, w_t was white Gaussian noise with a variance of 0.16. This is provided for completeness and not required by the algorithm (e.g. the noise could be biased and/or coloured). The input u_t was white Gaussian with variance of 4. The simulation was run for $t = 1, \ldots, N$ where the total number of data points was $N = 1000$. The parameters a^0_t and b^0_t varied linearly according to

$$a^0_t = \frac{-0.2}{N}t - 0.5$$
$$b^0_t = \frac{-0.2}{N}t + 0.5$$

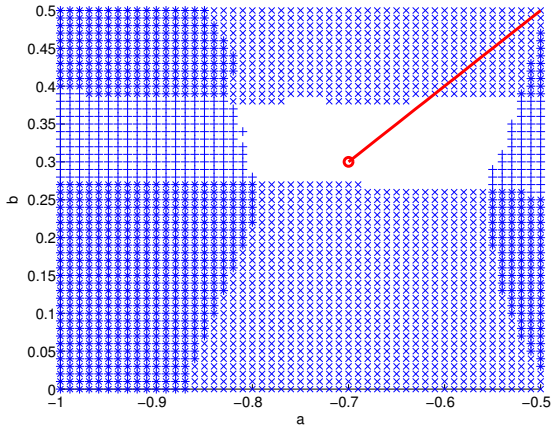

Fig. 1. An 90% confidence region for (a^0, b^0) at $t = 1000$ (blank region). The system trajectory is shown as the solid line with the final parameter indicated by $\circ$. $\lambda = 0.99$, $\boldsymbol{K} = [3 \times 10^{-4} \quad 3 \times 10^{-4}]^T$.

Thus the absolute value of the actual parameter change at each time step was

$$\boldsymbol{K}^0 = \begin{bmatrix} |\Delta a^0| \\ |\Delta b^0| \end{bmatrix} = \begin{bmatrix} 2 \times 10^{-4} \\ 2 \times 10^{-4} \end{bmatrix}$$

In this simulation we assumed the parameter change was 50% larger than the actual change. That is, $\boldsymbol{K} = [3 \times 10^{-4} \quad 3 \times 10^{-4}]^T$.

Intergers $q = 3$ and $M = 120$ were selected such that the probability the region contained the true parameter was $1 - \frac{2nq}{M} = 1 - \frac{2\times2\times3}{120} = 0.9$. A forgetting factor $\lambda = 0.99$ was chosen and the confidence set was calculated at each time step.

Figure 1 shows the system trajectory and the 90% confidence region at $t = 1000$. Figure 2 represents a one dimensional confidence interval for the parameters a^0_t and b^0_t by considering the width and height of the confidence region at the true parameter location. Time instants where the true parameter value did not belong to the confidence region are marked with an $\times$ at the bottom of the plot.

In this simulation, the parameter was always changing significantly less than the bound assumed by the algorithm. By selecting a bound equal to the actual change, that is $\boldsymbol{K} = [2 \times 10^{-4} \quad 2 \times 10^{-4}]^T$, the confidence set was, as expected, reduced as shown in Figure 3. The interval at each time was also smaller, and there were more times where the confidence set did not contain the true parameter as expected. See Figure 4.

The forgetting factor, λ, is a tradeoff between noise suppression and fast tracking of the parameters. The "effective" number of past data points used in the algorithm is $\frac{1}{1-\lambda}$ (see Ljung and Söderström [1983]) and hence the number of "effective" data points increases with λ. With $\lambda = 0.995$, the algorithm uses approximately 200 data points to calculate the confidence set. Figure 5 shows the confidence region at $t = 1000$, with $\lambda = 0.995$. As expected, due to the larger number of "effective" data points, the confidence region seems biased towards past parameter values resulting in a larger confidence region.

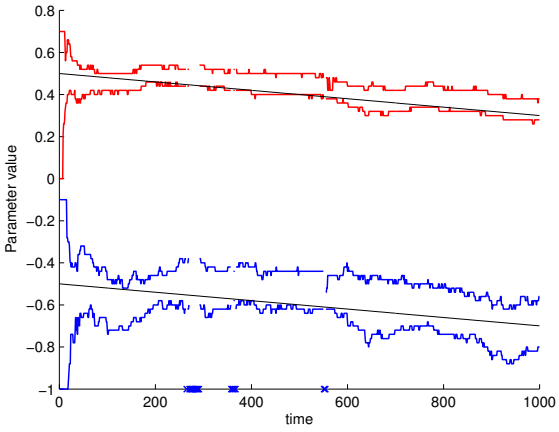

Fig. 2. The true parameter trajectories and 90% confidence region for a^0 (bottom lines) and b^0 (top lines) as time progresses. Time instants where the region doesn't include the true parameter are marked with an $\times$. $\lambda = 0.99$, $\boldsymbol{K} = [3 \times 10^{-4} \quad 3 \times 10^{-4}]^T$.

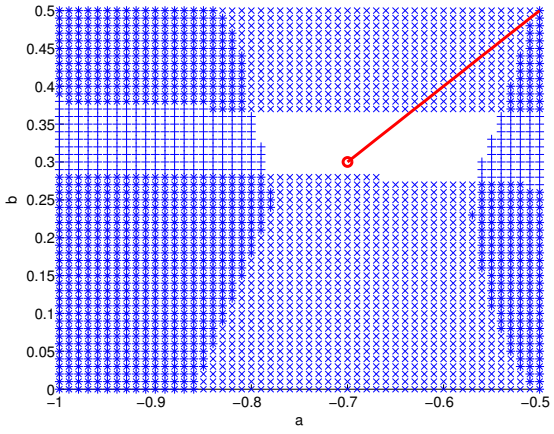

Fig. 3. An 90% confidence region for (a^0, b^0) at $t = 1000$ (blank region). The system trajectory is shown as the solid line with the final parameter indicated by $\circ$. $\lambda = 0.99$, $\boldsymbol{K} = [2 \times 10^{-4} \quad 2 \times 10^{-4}]^T$.

A larger λ also means that the region changes slower as demonstrated by Figure 6.

Alternatively, when λ is too small, the interval becomes noisy and fluctuates rapidly between time points. The case when $\lambda = 0.95$ is equivalent to using just 20 data points to calculate the confidence set. This situation was simulated and the result is shown in Figure 7.

5. CONCLUSION

In this paper, we have derived an algorithm for tracking the parameters of time-varying systems by assuming that an upper bound on the parameter variation is available. The algorithm delivers guaranteed confidence regions without any prior knowledge of the noise.

REFERENCES

M.C. Campi and E. Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005.

M.C. Campi and E. Weyer. Non-asymptotic confidence sets for input-output transfer functions. *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 157–162, December 2006a.

M.C. Campi and E. Weyer. Identification with finitely many data points: The LSCR approach. *Proceedings of the 14th IFAC Symposium on System Identification*, pages 46–64, March 2006b.

L. Ljung. *System Identification - Theory for the User*. Prentice Hall, second edition, 1999.

L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, 1983.

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

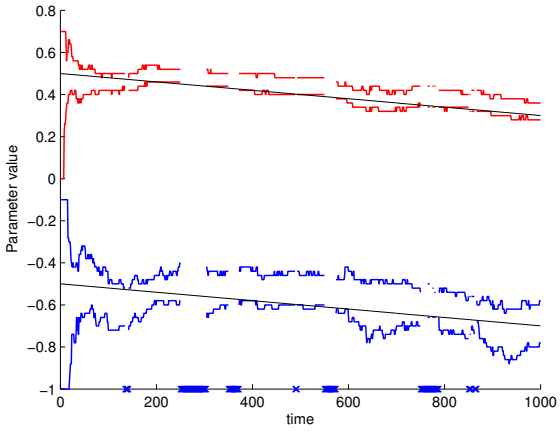

Fig. 4. The true parameter trajectories and 90% confidence region for a^0 (bottom lines) and b^0 (top lines) as time progresses. Regions that do not include the true parameter are marked with an $\times$. $\lambda = 0.99$, $\boldsymbol{K} = [2 \times 10^{-4} \quad 2 \times 10^{-4}]^T$.

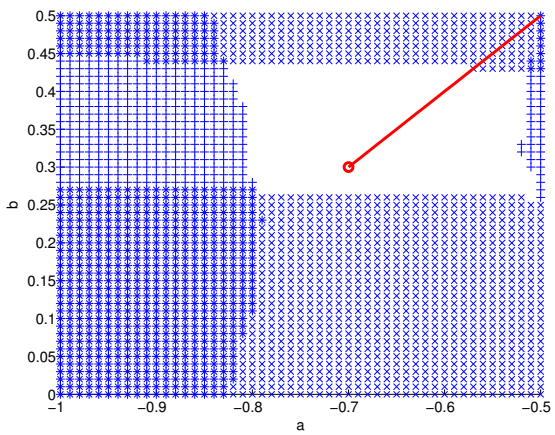

Fig. 5. An 90% confidence region for (a^0, b^0) at $t = 1000$ (blank region). The system trajectory is shown as the solid line with the final parameter indicated by $\circ$. $\lambda = 0.995$, $\boldsymbol{K} = [3 \times 10^{-4} \quad 3 \times 10^{-4}]^T$.

Appendix A. ALGORITHM DERIVATION

The term $g^k_{i,N}(\boldsymbol{\theta})$ in (7), can be computed as follows.

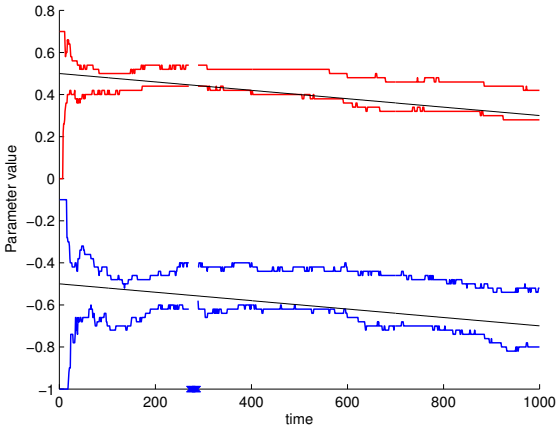

Fig. 6. The true parameter trajectories and 90% confidence region for a^0 (bottom lines) and b^0 (top lines) as time progresses. Regions that do not include the true parameter are marked with an $\times$. $\lambda = 0.995$, $\boldsymbol{K} = [3\times 10^{-4} \quad 3\times 10^{-4}]^T$.

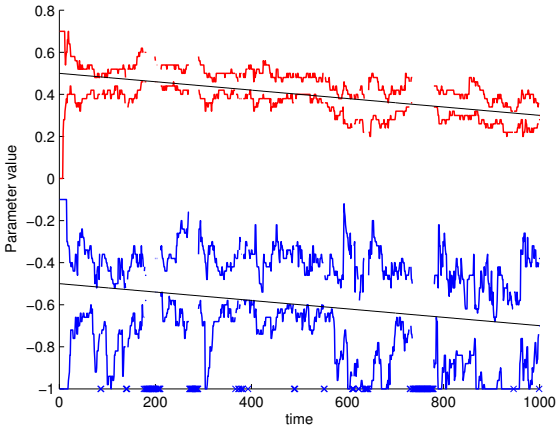

Fig. 7. The true parameter trajectories and 90% confidence region for a^0 (bottom lines) and b^0 (top lines) as time progresses. Regions that do not include the true parameter are marked with an $\times$. $\lambda = 0.95$, $\boldsymbol{K} = [3\times 10^{-4} \quad 3\times 10^{-4}]^T$.

$$
\begin{aligned}
g^k_{i,N}(\boldsymbol{\theta}) &= \sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\varepsilon_t(\boldsymbol{\theta})u_{t-k} \\
g^k_{i,N+1}(\boldsymbol{\theta}) &= \sum_{t=1}^{N+1} h_{i,t}\lambda^{N+1-t}\varepsilon_t(\boldsymbol{\theta})u_{t-k} \\
&= \lambda\sum_{t=1}^{N} h_{i,t}\lambda^{N-t}\varepsilon_t(\boldsymbol{\theta})u_{t-k} + \\
&\quad h_{i,N+1}\varepsilon_{N+1}(\boldsymbol{\theta})u_{N+1-k} \\
&= \lambda g^k_{i,N}(\boldsymbol{\theta}) + h_{i,N+1}\varepsilon_{N+1}(\boldsymbol{\theta})u_{N+1-k} \qquad \text{(A.1)}
\end{aligned}
$$

The term $C^k_{i,N}$ in (8), can be computed as follows

$$
\begin{aligned}
C^k_{i,N} &= \sum_{t=1}^{N-1} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t^T \left(\sum_{j=t}^{N-1}(\boldsymbol{\theta}^0_{j+1}-\boldsymbol{\theta}^0_j)\right) u_{t-k} \\
&= \sum_{j=1}^{N-1}(\boldsymbol{\theta}^0_{j+1}-\boldsymbol{\theta}^0_j)^T \sum_{t=1}^{j} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t u_{t-k} \\
C^k_{i,N+1} &= \sum_{j=1}^{N}(\boldsymbol{\theta}^0_{j+1}-\boldsymbol{\theta}^0_j)^T \sum_{t=1}^{j} h_{i,t}\lambda^{N+1-t}\boldsymbol{\phi}_t u_{t-k} \\
&= \lambda\sum_{j=1}^{N-1}(\boldsymbol{\theta}^0_{j+1}-\boldsymbol{\theta}^0_j)^T \sum_{t=1}^{j} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t u_{t-k} + \\
&\quad (\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N)^T \sum_{t=1}^{N} h_{i,t}\lambda^{N+1-t}\boldsymbol{\phi}_t u_{t-k} \\
&= \lambda C^k_{i,N} + (\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N)^T \boldsymbol{D}^k_{i,N+1} \\
&= \lambda C^k_{i,N} + {\boldsymbol{D}^k_{i,N+1}}^T(\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{D}^k_{i,N+1} &= \sum_{t=1}^{N} h_{i,t}\lambda^{N+1-t}\boldsymbol{\phi}_t u_{t-k} \\
&= \lambda\sum_{t=1}^{N-1} h_{i,t}\lambda^{N-t}\boldsymbol{\phi}_t u_{t-k} + h_{i,N}\lambda\boldsymbol{\phi}_N u_{N-k} \\
&= \lambda\boldsymbol{D}^k_{i,N} + \lambda h_{i,N}\boldsymbol{\phi}_N u_{N-k}
\end{aligned}
$$

The maximum possible value of $C^k_{i,N+1}$ at any given time is determined from the largest value of the product ${\boldsymbol{D}^k_{i,N+1}}^T(\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N)$. We know that $|\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N| \leq \boldsymbol{K}$ and the elements can have either sign to maximise the product. Thus

$$
|{\boldsymbol{D}^k_{i,N+1}}^T(\boldsymbol{\theta}^0_{N+1}-\boldsymbol{\theta}^0_N)| \leq |\boldsymbol{D}^k_{i,N+1}|^T\boldsymbol{K}
$$

For these new terms, a bound on $C^k_{i,N+1}$ is given by

$$
|C^k_{i,N+1}| \leq \bar{C}_{i,N+1}
$$

where

$$
\bar{C}^k_{i,N+1} = \lambda\bar{C}^k_{i,N} + |\boldsymbol{D}^k_{i,N+1}|^T\boldsymbol{K} \qquad \text{(A.2)}
$$

Thus corresponding to the true value $\boldsymbol{\theta} = \boldsymbol{\theta}^0_{N+1}$

$$
g^k_{i,N}(\boldsymbol{\theta}^0_{N+1}) - \bar{C}^k_{i,N+1} \leq f^k_{i,N+1} \leq g^k_{i,N+1}(\boldsymbol{\theta}^0_{N+1}) + \bar{C}^k_{i,N+1}
$$

where $f^k_{i,N+1}$ is given in (6).

Let

$$
\begin{aligned}
\boldsymbol{g}_{i,N}(\boldsymbol{\theta}) &= [g^1_{i,N}(\boldsymbol{\theta}),\ldots,g^n_{i,N}(\boldsymbol{\theta})]^T \\
\bar{\boldsymbol{C}}_{i,N} &= [\bar{C}^1_{i,N},\ldots,\bar{C}^n_{i,N}]^T \\
\boldsymbol{D}_{i,N} &= [\boldsymbol{D}^1_{i,N},\ldots,\boldsymbol{D}^n_{i,N}]^T
\end{aligned}
$$

It follows that (A.1) and (A.2) can be written in vector notation as follows

$$
\begin{aligned}
\boldsymbol{g}_{i,N+1}(\boldsymbol{\theta}) &= \lambda\boldsymbol{g}_{i,N}(\boldsymbol{\theta}) + h_{i,N+1}\varepsilon_{N+1}(\boldsymbol{\theta})\boldsymbol{\xi}_{N+1} \\
\boldsymbol{D}_{i,N+1} &= \lambda\boldsymbol{D}_{i,N} + \lambda h_{i,N}\boldsymbol{\phi}_N\boldsymbol{\xi}^T_{N+1} \\
\bar{\boldsymbol{C}}_{i,N+1} &= \lambda\bar{\boldsymbol{C}}_{i,N} + |\boldsymbol{D}_{i,N+1}|^T\boldsymbol{K}
\end{aligned}
$$