2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Prediction, filtering and smoothing using LSCR: State estimation algorithms with guaranteed confidence sets

Erik Weyer and Marco C. Campi

Abstract— **In this paper we present a new approach to state estimation based on the Leave-out Sign-dominant Correlation Regions (LSCR) principle from system identification. The LSCR approach constructs a confidence set for the state which has the property that, regardless of the value of the true state, the constructed confidence set will contain it with a guaranteed user chosen probability. This property makes the approach potentially very useful in monitoring and detection applications. The proposed approach is illustrated in simulation examples.**

I. Introduction

We consider state estimation of a stable, linear discrete time system

$$x_{t+1} = Fx_t + Gv_t, \quad y_t = Hx_t + w_t$$

where $x_t \in R^n$ and y_t is a scalar. The noise processes v_t and w_t are scalar mutually independent white Gaussian noise sequences. F, G and H are matrices of appropriate dimensions. If the variances σ_v^2 and σ_w^2 and the initial state x_0 are Gaussian, then the optimal minimum variance solution is given by the Kalman filter (e.g. [1]). Here we consider the case where σ_v^2 and σ_w^2 are unknown, and we show that by using the backward representation of the process x_t ([8]) and the Karhunen-Loeve transformation (e.g. [9]), this problem can be approached using the LSCR algorithm from system identification ([2], [3]). The LSCR algorithm delivers a confidence set which has a user chosen guaranteed probability of containing the true state vector.

The nature of the statements about the true vector in our results obtained using LSCR are different from the ones for the Kalman filter. Using the Kalman filter one obtains a Gaussian state estimate and a covariance matrix. Based on the state estimate and the covariance matrix one can construct a confidence ellipsoid and make a statement that the true state belongs to this confidence ellipsoid with a given probability. The LSCR result statement is different as it says, *regardless* of what the true state is, the constructed confidence set will contain the true state with a given probability. This type of result is more relevant for monitoring applications, where we are much more concerned about obtaining reliable confidence sets when the state takes on particular values, e.g. when the state is in or close to a dangerous operating region.

The work of E. Weyer was supported by the Australian Research Council under Discovery Grant DP0986162.

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

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

The reason we can obtain this stronger statement is that we condition with respect to the state to be estimated such that it can be treated as a deterministic variable in the LSCR framework. As the observed variables in the LSCR framework needs to be conditionally independent, the backward representation of x_t is employed for the prediction and filtering problem, and a combination of the backward and forward representation is used for the smoothing problem.

A number of other approaches to filtering under uncertainty can be found in the literature, e.g. [11], [13], [5], [6], [4], [12], [10]. An often used problem setting is that the system matrices and/or noise covariance matrices belongs to certain bounded sets, and a filter which in some sense minimises and upper bound on the estimation error is derived.

The paper is organised as follows. In the next section we introduce the backward representation of a state space system. The prediction problem is treated in detail in Section III, while the filtering and smoothing problems are briefly discussed in Section IV. Simulation examples are given in Section V followed by some concluding remarks.

II. Backward and forward representation of linear systems.

A. Forward representation

Consider the standard (forward) model

$$x_{t+1} = F^f x_t + G^f v_t \quad (1)$$

$$y_t = Hx_t + w_t \quad (2)$$

where $x_t \in R^n$, $y_t \in R$, $F^f \in R^{n\times n}$ and $H \in R^{1\times n}$. v_t and w_t are scalar noise processes and hence $G^f \in R^{n\times 1}$. We make the following assumptions

Assumptions

1) v_t and w_t are mutually independent sequences of independent and identically distributed Gaussian random variables with unknown variances σ_v^2 and σ_w^2 respectively.
2) F^f has all eigenvalues strictly inside the unit circle, and x_t is a stationary process.

The covariance matrix Π of x_t is given by the equation

$$\begin{aligned}\Pi &= F^f \Pi F^{fT} + G^f G^{fT}\sigma_v^2 \\ &= F^f(F^f \Pi F^{fT} + G^f G^{fT}\sigma_v^2)F^{fT} + G^f G^{fT}\sigma_v^2 \\ &= \sigma_v^2(G^f G^{fT} + F^f G^f G^{fT} F^{fT} + F^{f2} G^f G^{fT} F^{f2T} + \cdots) \\ &= \sigma_v^2 \Pi'\end{aligned} \quad (3)$$

The expansion on the third line converges since F^f has all eigenvalues inside the unit circle, and Π' is given by the equation $\Pi' = F^f \Pi' F^{fT} + G^f G^{fT}$.

B. Backward representation

The process x_t can also be represented by a state equation running backwards in time (see e.g. Lemma 5.4.4 in [8])

$$\begin{aligned} x_t &= F^b x_{t+1} + v_t^b && (4)\\ y_t &= H x_t + w_t && (5)\end{aligned}$$

where F^b is a solution to the equation $F^b\Pi = \Pi F^{fT}$. For simplicity we assume that Π is non-singular such that

$$F^b = \Pi F^{fT}\Pi^{-1} = \Pi' F^{fT}\Pi'^{-1} \quad (6)$$

which does not depend on the value of σ_v^2. Note that while v_t in (1) is scalar, v_t^b in (4) is an independent and identically distributed Gaussian *vector*-process with covariance matrix Q independent of the process w_t. Q is given by

$$Q = \Pi - F^f \Pi F^{fT} = \sigma_v^2(\Pi' - F^f \Pi' F^{fT}) = \sigma_v^2 Q' \quad (7)$$

where $Q' = \Pi' - F^f \Pi' F^{fT}$, hence Q is known up to the scalar factor σ_v^2.

C. Combination of forward and backward representation

The forward and backward representations can be combined such that the process x_t is represented by the backward representation up to time $k-1$ and by the forward representation from time $k+1$ with the state x_k acting as the initial condition, i.e.

$$\begin{aligned} x_{t-1} &= F^b x_t + v_{t-1}^b,\ t = k, k-1, \dots && (8)\\ x_{t+1} &= F^f x_t + G^f v_t,\ t = k, k+1, \dots && (9)\\ y_t &= H x_t + w_t && (10)\end{aligned}$$

We have the following results

Lemma 2.1: $v_t, \dots, v_k, v_{k-1}^b, \dots, v_1^b$ and x_k are mutually independent Gaussian random variables.

Proof. See [8], Ch. 10.4. □

III. Prediction

Here the problem of estimating x_{t+1}, given observations $y_1, \dots, y_t$ is considered. Using the backward representation we have that

$$y_t = H x_t + w_t = H F^b x_{t+1} + H v_t^b + w_t$$

and

$$y_{t-1} = H F^{b2} x_{t+1} + H F^b v_t^b + H v_{t-1}^b + w_{t-1}$$

Hence, by repeated substitution the observed data can be expressed in terms of the state x_{t+1} and the noise processes v^b and w. Let

$$B = \begin{bmatrix} H & 0 & 0 & \cdots & 0 \\ HF^b & H & 0 & \cdots & 0 \\ HF^{b2} & HF^b & H & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ HF^{b(t-1)} & HF^{b(t-2)} & HF^{b(t-3)} & \cdots & H \end{bmatrix}$$

then we obtain the following set of equations

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} = \begin{bmatrix} HF^b \\ HF^{b2} \\ HF^{b3} \\ \vdots \\ HF^{bt} \end{bmatrix} x_{t+1} + B \begin{bmatrix} v_t^b \\ v_{t-1}^b \\ v_{t-2}^b \\ \vdots \\ v_1^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix} \quad (11)$$

The covariance matrix of the second term on the right hand side of (11) is given by

$$\begin{aligned} Q_{v^b} &= \sigma_v^2 B \begin{bmatrix} Q' & 0 & \cdots & 0 \\ 0 & Q' & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & & Q' \end{bmatrix} B^T \\ &= \sigma_v^2 Q'_{v^b} = \sigma_v^2 U D U^T && (12)\end{aligned}$$

where UDU^T is the singular value decomposition of the known matrix Q'_{v^b}. By premultiplying (11) with U^T the following set of equations is obtained

$$\begin{bmatrix} s_t \\ s_{t-1} \\ s_{t-2} \\ \vdots \\ s_1 \end{bmatrix} = \begin{bmatrix} a_t \\ a_{t-1} \\ a_{t-2} \\ \vdots \\ a_1 \end{bmatrix} x_{t+1} + \begin{bmatrix} n_t \\ n_{t-1} \\ n_{t-2} \\ \vdots \\ n_1 \end{bmatrix} \quad (13)$$

or more compactly $s = A x_{t+1} + \bar{n}$ where

$$s = \begin{bmatrix} s_t \\ s_{t-1} \\ s_{t-2} \\ \vdots \\ s_1 \end{bmatrix} = U^T \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} \quad (14)$$

$$A = \begin{bmatrix} a_t \\ a_{t-1} \\ a_{t-2} \\ \vdots \\ a_1 \end{bmatrix} = U^T \begin{bmatrix} HF^b \\ HF^{b2} \\ HF^{b3} \\ \vdots \\ HF^{bt} \end{bmatrix} \quad (15)$$

and

$$\bar{n} = \begin{bmatrix} n_t \\ n_{t-1} \\ n_{t-2} \\ \vdots \\ n_1 \end{bmatrix} = U^T B \begin{bmatrix} v_t^b \\ v_{t-1}^b \\ v_{t-2}^b \\ \vdots \\ v_1^b \end{bmatrix} + U^T \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix}$$

Note that while the components s_k and n_k of s and $\bar{n}$ are scalars, the components a_k of A are $1 \times n$ vectors. The components of $\bar{n}$ are independent of each other as shown in the next Lemma.

Lemma 3.1: $n_t, n_{t-1}, \dots, n_1$ are mutually independent zero mean Gaussian random variables.

Proof. The covariance matrix of $\bar{n} = [n_t, \dots, n_1]^T$ is given by the diagonal matrix $\sigma_v^2 D + \sigma_w^2 I$. As both v^b and w are Gaussian processes with zero mean, $n_t, \dots n_1$ are

mutually independent (but not identically distributed) zero mean Gaussian variables. □

We are now in a position that we can construct a confidence set for x_{t+1} using the LSCR algorithm ([2], [3]). Let x denote a generic value of x_{t+1}, and let $\hat{s}_t(x) = a_t x$ be a prediction of s_t. Furthermore, let the singular value decomposition of the $t \times n$ matrix A be given by

$$U_A \Sigma V_A^T = A \tag{16}$$

where U_A is a $t \times t$ matrix, Σ a $t \times n$ matrix and V_A an $n \times n$ matrix. Finally, let $[u_{a,l,t}, u_{a,l,t-1}, \dots u_{a,l,1}]^T$ denote the lth column of U_A. The confidence set of x_{t+1} can now be constructed as follows

Construction of confidence sets. Given: Observations $y_1, \dots, y_t$ and system matrices F^f, G^f and H as in (1)-(2).

1) Using the given data and the system matrices, compute s, A and U_A from equations (14), (15) and (16) using (3), (6), (7) and (12).
2) Compute the prediction errors
$$\epsilon_k(x) = s_k - a_k x \quad k = 1, \dots, t.$$
3) For $l = 1, \dots, n$ and $k = 1, \dots, t$ compute the correlation
$$f_{k,l}(x) = \epsilon_k(x) u_{a,l,k} = (s_k - a_k x) u_{a,l,k}.$$
4) Select an integer M and construct M binary ($\{0,1\}$-valued) stochastic strings of length t as follows: Let $h_{0,1}, \dots, h_{0,t} = 0, \dots, 0$ 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. However, if a string turns out to be equal to an already constructed string, this string is removed and another string to be used in its place is constructed according to the same rule. Name the constructed non-zero strings $h_{1,1}, \dots, h_{1,t}$; $h_{2,1}, \dots, h_{2,t}$;...; $h_{M-1,1}, \dots, h_{M-1,t}$. For $l = 1, \dots, n$ and $i = 0, \dots, M-1$ compute
$$\begin{aligned} g_{i,l}(x) &= \sum_{k=1}^{t} h_{i,k} \cdot f_{k,l}(x) \\ &= \sum_{k=1}^{t} h_{i,k} \cdot (s_k - a_k x) u_{a,l,k}. \end{aligned}$$
Note that $g_0(x) \equiv 0$.
5) Select an integer q in the interval $[1, M/2n]$. Let the sets $\hat{X}_l$, $l = 1, \dots, n$ contain those values of x such that at least q of the $g_{i,l}(x)$ functions ($i = 0, \dots, M-1$) are strictly larger than $g_{0,l}(x) \equiv 0$ and at least q are strictly smaller than $g_{0,l}(x) \equiv 0$.
6) The confidence set is given by
$$\hat{X} = \cap_{l=1}^{n} \hat{X}_l. \tag{17}$$

The constructed confidence sets have the following properties.

Theorem 3.2: Consider the system given by (1)-(2) and assume that assumptions 1 and 2 in Section II-A are satisfied. Then, regardless of what the true value of x_{t+1} is, the sets $\hat{X}_l$ and $\hat{X}$ constructed above have the properties

$$Pr\{x_{t+1} \in \hat{X}_l\} = 1 - 2q/M, \tag{18}$$
$$Pr\{x_{t+1} \in \hat{X}\} \geq 1 - 2qn/M, \tag{19}$$

where q and M are introduced in points 5 and 4 respectively.

Proof. See proof of Theorem 1 in [3]. □

Remark. Note that corresponding to $x = x_{t+1}$ (the true value), $g_{i,l}(x_{t+1}) = \sum_{k=1}^{t} h_{i,k} u_{a,l,k} n_k$, $i = 0, \dots, M-1$ is a sum of independent random variables symmetrically distributed around 0. It is therefore unlikely that nearly all of the M sums are positive or negative corresponding to $x = x_{t+1}$ and those values of x for which this happens are excluded from the confidence set. □

Remark. A crucial point in the proof of Theorem 3.2 is that $u_{a,l,k} n_k$ is a sequence of independent (but not necessarily identically distributed) random variables symmetric about 0. This has been achieved via two important steps.

1) The premultiplication with U^T in (13). U^T has the property that it diagonalises both the covariance matrices Q_{v^b} and $\sigma_w^2 I$, that is $U^T Q_{v^b} U = \sigma_v^2 D$ and $U^T \sigma_w^2 I U = \sigma_w^2 I$, and since the processes are Gaussian this means that the corresponding random variables are independent Gaussian. This diagonalisation is the same as applying the Karhunen-Loeve decomposition for decorrelation of a Gaussian stochastic process (e.g. [9]).
2) The use of the backward representation. Using the backward representation allows us to condition on x_{t+1}, since both the noise sequences $v_t^b, \dots, v_1^b$ and $w_t, \dots, w_1$ are conditionally independent of x_{t+1}. Note that in the forward representation the noise sequence $v_1, \dots, v_t$ is not an iid sequence given x_{t+1}. Moreover, it is the conditioning with respect to x_{t+1} that leads to the conclusion that the true state is in the confidence set with a certain probability *regardless* of the value of the true state as discussed in the next subsection.

□

Remark. An alternative to using the stochastic strings in point 4 in the algorithm above is to use the following construction based upon groups under the symmetric difference operation

4a. Let $I = \{1, \dots, t\}$ and consider a collection G of subsets $I_i \subseteq I$, $i = 1, \dots, M$, with $I_0 = \emptyset$ forming a group under the symmetric difference operation (i.e. $(I_i \cup I_j) - (I_i \cap I_j) \in G$, if $I_i, I_j \in G$). Compute
$$g_{i,l}(x) = \sum_{k \in I_i} (s_k - a_k x) u_{a,l,k}, \quad i = 1, \dots, M;$$
with $g_{0,l}(x) \equiv 0$.

Theorem 3.2 holds unaltered with this construction (see [2]). This construction can be an advantage when there are few data points. Well suited groups for the construction of confidence sets are given in [7]. □

A. Interpretation

Note that the theorem statement above is conceptually different from what a Kalman filter would deliver. At time t the above theorem states that : "Whatever the next value x_{t+1} is, the set $\hat{X}$ will contain it with probability at least $1 - 2qn/M$". Using a Kalman filter based on (1)-(2) and assuming the noise variances σ_v^2 and σ_w^2 are known, one would obtain an a-priori estimate $\hat{x}_{t+1|t}$ with a Gaussian distribution and a covariance matrix given by the Riccati equation. From the a-priori estimate and the covariance matrix one can construct a confidence ellipsoid $\hat{X}_{\text{KF}}$ which has the interpretation "the next value x_{t+1} is in the confidence ellipsoid $\hat{X}_{\text{KF}}$ with probability $1 - p$".

For certain problems the statement in Theorem 3.2 is more relevant. One example is a monitoring problem. Say that the state should not enter a critical region and that the probability for the state to enter the region is very low. In principle whenever the state enters the region the confidence ellipsoid generated by the Kalman filter may never contain it as it would be one of those events with probability p for which the confidence ellipsoids do not contain the true value. The statement in Theorem 3.2 is stronger as it says, whatever the true value is, it will be contained in the generated confidence set with a certain probability, and hence even if the state is in the undesired region the confidence set generated by the LSCR algorithm will contain it with the given probability.

Note that if the variances σ_v^2 and σ_w^2 where known, the covariance matrix of $\bar{n}$ in (13) would be known and the best linear unbiased estimate (BLUE) (also called minimum variance unbiased estimate) (e.g. [8]) of x_{t+1} could be computed together with its covariance matrix and the corresponding confidence ellipsoid. Like the result in Theorem 3.2, the BLUE confidence ellipsoid will contain the true x_{t+1} with a certain probability regardless of its value.

Note that in order for the confidence set $\hat{X}$ to be bounded and the BLUE to be unique, the matrix A in (15) must have full column rank, which is equivalent to $[(HF^b)^T, (HF^{b2})^T, \ldots, (HF^{bt})^T]^T$ has full column rank. Assuming $t \geq n$ (the dimension of x_t), this is guaranteed if F^b is non-singular and the system is observable.

IV. Filtering and smoothing

A. Filtering

In filtering one wants to estimate x_t based on measurements $y_1, \ldots, y_t$. It is treated in the same way as prediction. The only difference is that equation (11) takes the form

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{bmatrix} = \begin{bmatrix} H \\ HF^b \\ HF^{b2} \\ \vdots \\ HF^{b(t-1)} \end{bmatrix} x_t + B' \begin{bmatrix} v_{t-1}^b \\ v_{t-2}^b \\ v_{t-3}^b \\ \vdots \\ v_1^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_1 \end{bmatrix} \tag{20}$$

where the first row of B' is all 0s and the other rows are rows of B (as a $(t-1) \times (t-1)$) matrix shifted one row down. The rest of the development follows as for prediction.

B. Smoothing

Here we want to estimate x_i, $1 \leq i < t$ based upon $y_1, \ldots, y_t$. Let

$$B_1 = \begin{bmatrix} HG & HF^fG & \cdots & HF^{f(t-1-i)}G \\ 0 & HG & \cdots & HF^{f(t-2-i)}G \\ & & \ddots & \vdots \\ & & & HG \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} H & 0 & \cdots & 0 \\ HF^b & H & & \\ \vdots & \vdots & \ddots & \\ HF^{b(i-2)} & HF^{b(i-3)} & \cdots & H \end{bmatrix}$$

$$B'' = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}$$

Making use of both the forward and backward representation $[y_t, y_{t-1}, \ldots, y_{i+1}, y_i, y_{i-1}, y_{i-2}, \ldots, y_1]^T$ can be expressed as

$$\begin{bmatrix} HF^{f(t-i)} \\ HF^{f(t-1-i)} \\ \vdots \\ HF^f \\ H \\ HF^b \\ HF^{b2} \\ \vdots \\ HF^{b(i-1)} \end{bmatrix} x_i + B'' \begin{bmatrix} v_{t-1} \\ v_{t-2} \\ \vdots \\ v_i \\ v_{i-1}^b \\ v_{i-2}^b \\ \vdots \\ v_1^b \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t-1} \\ \vdots \\ w_{i+1} \\ w_i \\ w_{i-1} \\ w_{i-2} \\ \vdots \\ w_1 \end{bmatrix} \tag{21}$$

Recalling Lemma 2.1, $v_{t-1}, \ldots v_i, v_{i-1}^b, \ldots, v_1^b$ and x_i are mutually independent, and we can proceed as in the prediction case.

V. Simulation examples

A. Second order system

Here we illustrate the LSCR approach by applying it to a prediction and a smoothing problem and comparing the obtained confidence sets to the confidence ellipsoids obtained using the BLUE estimator. Consider a cascade of two first order systems driven by white noise with state space representation

$$x_{t+1} = \begin{bmatrix} 0.5 & 0 \\ 1 & 0.7 \end{bmatrix} x_t + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t, \quad y_t = [0\ 1] x_t + w_t$$

v_t and w_t are mutually independent white Gaussian sequences with unknown variances σ_v^2 and σ_w^2.

The backward representation is given by

$$x_t = \begin{bmatrix} 0.2450 & 0.3315 \\ -0.3500 & 0.9550 \end{bmatrix} x_{t+1} + v_t^b$$

where v_t^b is white Gaussian with covariance matrix

$$Q = \sigma_v^2 \begin{bmatrix} 0.4900 & -0.7000 \\ -0.7000 & 1.0000 \end{bmatrix}$$

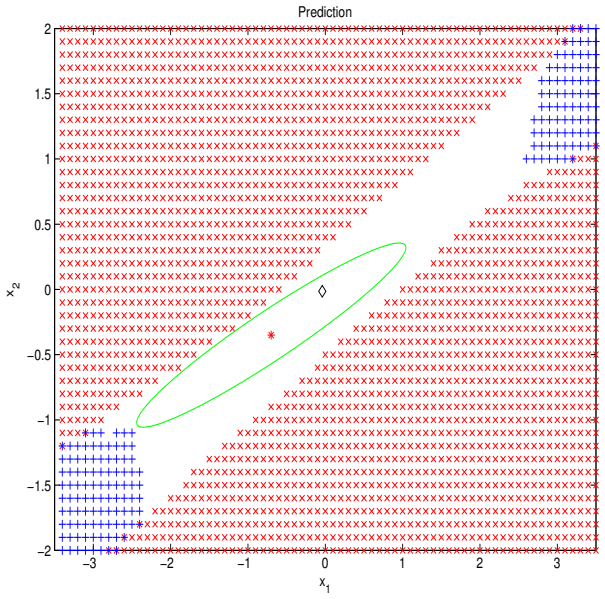

Fig. 1. Confidence set for x_{31}

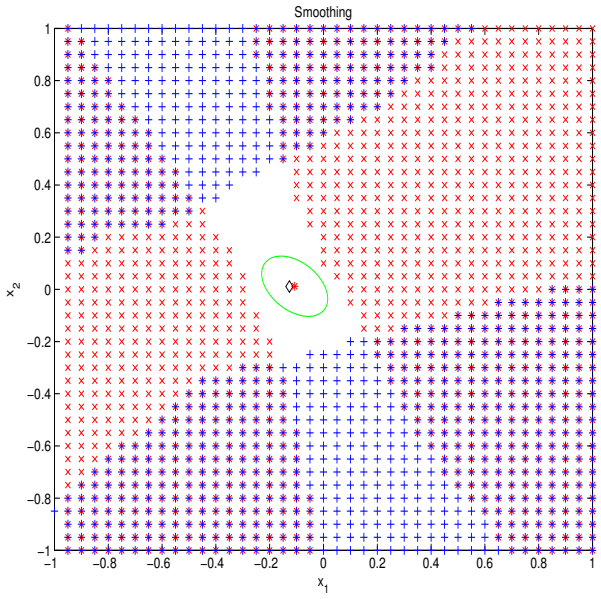

Fig. 2. Confidence set for x_{15}

We observe $t = 30$ data points $y_1, \ldots, y_{30}$ and we would like to predict the state x_{31} and to estimate the state x_{15}.

1) Prediction: The data are first transformed using (14) and (15). In order to construct a 90% confidence set for x_{31} we generated $M = 80$ stochastic strings as in point 4 of the algorithm in Section III. Next the correlation functions

$$g_{i,1}(x) = \sum_{k=1}^{30} h_{i,k}(s_k - a_k x) u_{a,1,k}, \quad i = 0, \ldots, 79$$

$$g_{i,2}(x) = \sum_{k=1}^{30} h_{i,k}(s_k - a_k x) u_{a,2,k}, \quad i = 0, \ldots, 79$$

were computed. The values of x for which 0 were among the $q = 4$ largest or smallest value of the two functions were excluded. The areas where 0 were among the 4 smallest or largest values of $g_{i,1}$ are marked with x in Fig. 1, and the corresponding areas for $g_{i,2}$ are marked with +. The obtained confidence region is the blank region in Fig. 1 and it is guaranteed to contain the true state x_{31} with probability at least $1 - 2 \cdot 2 \cdot 2/80 = 0.9$ regardless of the value of x_{31}. The true value is marked with a $\diamond$. Note that no knowledge about the noise variances were required nor used in the above procedure.

If the noise variances were known the BLUE estimate could have been computed. In this simulation $\sigma_v^2 = \sigma_w^2 = 0.005$, and the BLUE estimate is marked with a $\star$ and the corresponding 90% confidence ellipsoid is plotted. The confidence region obtained using the LSCR algorithm in Section III is similarly shaped to the confidence ellipsoid obtained from BLUE, but as expected it is larger since no knowledge of noise variances is assumed.

2) Smoothing: Transforming the data using equation (21) as a starting point, a 90% confidence set for the state x_{15} is obtained as above by computing 80 correlation functions using subsamples and discarding those values of the state vector for which zero were among the 4 smallest or largest values. The results are shown in Figure 2 together with the BLUE estimate and the corresponding 90% confidence ellipsoid. Not surprisingly these confidence sets are smaller than in the case of prediction. As for prediction the confidence region obtained using the LSCR algorithm is similarly shaped to the confidence ellipsoid obtained from BLUE, but larger since no knowledge of noise variances is assumed nor utilised.

B. Difference between LSCR prediction and Kalman filter prediction

Consider a first order moving average process

$$y_t = v_t + v_{t-1}$$

where v_t is white Gaussian noise with variance σ_v^2. This system has the state space representation

$$x_t = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} v_t, \quad y_t = [0\ 1] x_t \qquad (22)$$

Here

$$F^f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad G^f = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad H = [0\ 1]$$

There is no measurement noise so $\sigma_w^2 = 0$. Given observation up to time t we want to predict x_{t+1}. The Kalman filter is given by

$$\begin{aligned} K_t &= F^f \Sigma_t H^T (H \Sigma_t H^T)^{-1} \\ \Sigma_t &= F^f \Sigma_{t-1} F^{fT} + G^f G^{fT} \sigma_v^2 - \\ &\quad F^f \Sigma_{t-1} H^T (H \Sigma_{t-1} H^T)^{-1} H \Sigma_{t-1} F^{fT} \\ \hat{x}_{t|t-1} &= F^f \hat{x}_{t-1|t-2} + K_t (y_t - H \hat{x}_{t-1|t-2}) \end{aligned}$$

with initial conditions

$$\hat{x}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \sigma_v^2$$

K_t is the Kalman gain and Σ_t is the error covariance matrix $\Sigma_t = E[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T | y_{t-1}, \ldots, y_1]$.

Substituting the above values we find that

$$K_t = \begin{bmatrix} 0 \\ \frac{t}{t+1} \end{bmatrix}, \quad \Sigma_t = \begin{bmatrix} 1 & 1 \\ 1 & \frac{t+1}{t} \end{bmatrix} \sigma_v^2$$

The two elements of the state estimates are given by $\hat{x}_{t|t-1,1} = 0$ and $\hat{x}_{t|t-1,2} = ((t-1)y_{t-1} - (t-2)y_{t-2} + (t-3)y_{t-3} - \cdots + (-1)^t y_1)/t$.

The backward representation of the system (22) is

$$x_t = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} x_{t+1} + v_t^b, \quad y_t = [0 \ 1] x_t \qquad (23)$$

where $v_t^b = [0 \ v_{t-1}]$. Here F^b is singular. The first state is a white Gaussian noise process, and the Kalman filter estimate of it is zero and the variance is σ_v^2 regardless of the observations. This means that whenever $x_{1,t}$ is larger than a certain value it will not be in the confidence ellipsoid generated by the Kalman filter. In the backward representation the matrix F^b is singular which means that whenever x is in the confidence set generated by the LSCR algorithm so is $x + c[1, \ 1]^T$ for any real c, so it still holds true that the confidence set contains the true value of $x_{1,t}$ regardless of its value with a certain probability. In fact, irrespectively of the method used to construct the confidence set, since the first state is a white Gaussian noise process, the only way we can guarantee it is in the confidence set regardless of its value, is to have an unbounded confidence sets. In Fig. 3 the 90% confidence regions for x_{21} for the two methods are plotted together with the true value (diamond) and the Kalman filter estimate $\hat{x}_{21|20}$ (star). We note that the LSCR region (the blank area) is unbonded while the Kalman filter region is always centered at $x_{1,t} = 0$ and will not contain the true value of $x_{1,t}$ for large values of $x_{1,t}$. Also note that the Kalman filter requires knowledge of $\sigma_v^2 = 0.1$ and $\sigma_w^2 = 0$ while no such knowledge is required nor used in the LSCR approach. Moreover, unlike the Kalman filter the LSCR method is not a Bayesian approach since the conditioning with respect to the state to be estimated allows us to treat it as a deterministic variable.

VI. Conclusions

In this paper we have proposed a new approach to state estimation. By using the backward representation and the Karhunen-Loeve decomposition, the original system was transformed such that the LSCR algorithm from system identification could be applied. Using the LSCR algorithm, a confidence set was constructed which had the property that it contains the true state, regardless of its value, with a guaranteed user chosen probability. Moreover, no knowledge of the noise variances was required in order to construct the confidence set. The approach and how it differs from the Kalman filter were illustrated in simulation examples.

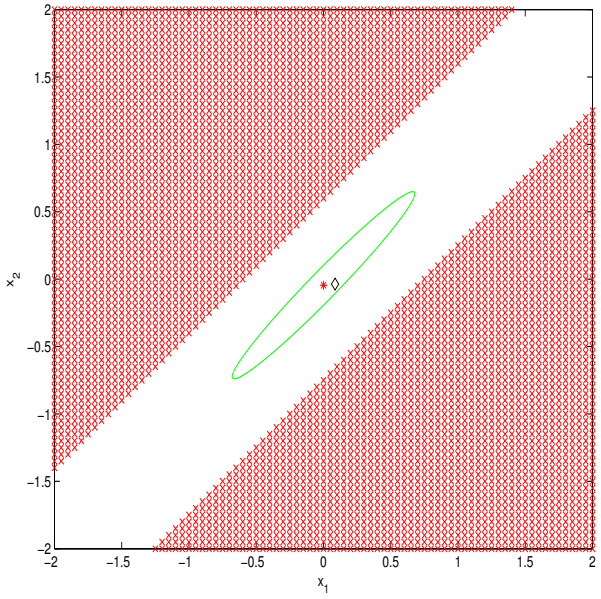

Fig. 3. Confidence sets for x_{20}. Blank region - LSCR. Ellipse - Kalman filter

References

[1] Anderson B.D.O. and J.B. Moore (1979). *Optimal Filtering.* Prentice Hall.

[2] Campi, M.C. and E. Weyer (2005). "Guaranteed non-asymptotic confidence regions in system identification", *Automatica* Vol. 41, no. 10, pp. 1751-1764.

[3] Campi, M.C. and E. Weyer (2010). "Non-asymptotic confidence sets for the parameters of linear transfer functions", *IEEE Transactions on Automatic Control*, Vol. 55, no. 12, pp. 2708-2720.

[4] Dong, Z. and Z. You (2006). "Finite-Horizon Robust Kalman Filtering for Uncertain Discrete Time-Varying Systems With Uncertain-Covariance White Noises", *IEEE Signal Processing Letters*, Vol. 13, no. 8, pp. 493-496.

[5] Fu, M., C.E. de Souza, and Z-Q Luo (2001). "Finite-Horizon Robust Kalman Filter Design" *IEEE Transactions on Signal Processing*, Vol. 49, no. 9, pp. 2103-2112.

[6] El Ghaoui, L. and G. Calafiore (2001). "Robust Filtering for Discrete-Time Systems with Bounded Noise and Parametric Uncertainty", *IEEE Transactions on Automatic Control*, Vol. 46, no. 7, pp. 1084-1089.

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

[8] Kailath, T., A.H. Sayed, and B. Hassibi (2000). *Linear estimation.* Prentice Hall.

[9] Levy B.C. (2008). *Principles of signal detection and parameter estimation*. Springer.

[10] Pankov, A.R., E.N. Platonov, and K.V. Semenikhin (2011). "Robust filtering of process in the stationary difference stochastic system", *Automation and Remote Control*, Vol. 72, no. 2, pp. 377-392.

[11] Mehra, R.K. (1970). "On the identification of variances and adaptive Kalman filtering", *IEEE Transactions on Automatic Control*, Vol. 15, no. 2, pp. 175-184.

[12] Shmaliy, Y. S. (2010). "Linear Optimal FIR Estimation of Discrete Time-Invariant State-Space Models", *IEEE Transactions on Signal Processing*, Vol. 58, no. 6, pp. 3086-3096.

[13] Theodor, Y. and U. Shaked (1996). "Robust Discrete-Time Minimum-Variance Filtering", *IEEE Transactions on Signal Processing*, Vol. 44, no. 2, pp. 181-189.