AVOIDING WINDUP IN RECURSIVE PARAMETER ESTIMATION

Benny Stenlund * **Fredrik Gustafsson** **

* *Control Engineering Group*
Luleå University of Technology
SE-971 87 Luleå, Sweden
Email:benny.stenlund@sm.luth.se
** *Dept. of Electrical Engineering*
Linköping University
SE-58183 Linköping, Sweden
Email:fredrik@isy.liu.se

Abstract: Recursive parameter estimation algorithms can be recast in a Kalman filter setting with different choices of parameter random walk covariance matrix Q. In the (unrealistic) case where the designed Q is equal to the true one, the Kalman filter covariance matrix P can be used as the covariance matrix of the parameter estimation error. However, often Q is chosen to get an appropriate time constant of the filter and P is only instrumental. A problem of practical interest is to avoid windup in P during periods of poor excitation, leading to uncertain estimates and long transients when the data becomes exciting. Here we survey different methods to avoid windup, with the goal to accept temporarily longer time constants in the directions of the parameter space where no or poor excitation appears. A low-dimensional example, motivated by a practically important application, illustrates the ideas.

Keywords: Recursive estimation, Adaptive filtering, Parameter estimation, Time-varying systems

1. INTRODUCTION

We consider signals generated by linear regression models

$$y(t) = \varphi(t)^T \theta(t) + e(t), \quad R^o = \mathbf{C}ov(e(t)). \tag{1}$$

The estimation model is assumed to be of the correct linear regression structure and a random walk model of the parameter variation:

$$\theta(t+1) = \theta(t) + v(t), \quad Q = \mathbf{C}ov(v(t)), \tag{2a}$$

$$y(t) = \varphi(t)^T \theta(t) + e(t), \quad R = \mathbf{C}ov(e(t)). \tag{2b}$$

The Kalman filter applied to this model gives

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\varepsilon(t) \tag{3a}$$

$$\varepsilon(t) = [y(t) - \varphi(t)^T \hat{\theta}(t-1)] \tag{3b}$$

$$K(t) = \frac{P(t-1)\varphi(t)}{R(t) + \varphi(t)^T P(t-1)\varphi(t)} \tag{3c}$$

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi(t)^T P(t-1)}{R(t) + \varphi(t)^T P(t)\varphi(t)} + Q(t). \tag{3d}$$

In the case $v(t)$ and $e(t)$ are white noise sequences with Gaussian distributions, the Kalman filter provides the best estimate in the sense of a minimal conditional estimation error covariance matrix. If $v(t)$ and $e(t)$ are non-gaussian the Kalman filter is still the best linear estimator, see e.g. (Anderson and Moore, 1979). One can easily show that the well-known recursive least squares

(RLS) and least means square (LMS) algorithm is included as special cases of the Kalman filter corresponding to specific assumptions about the covariance matrix $Q(t)$ of the parameter variations, see e.g. (Gunnarsson, 1991) and (Ljung and Gunnarsson, 1990).

The problem in using the Kalman filter as a tracking algorithm (which in practice means that Q is chosen as a constant positive definite matrix, not necessarily the true value) is windup of the matrix P when signals are poorly exciting. This means that some of the eigenvalues of P tends to unacceptable large values. One then gets numerical problems and high sensitivity against noise. The same problem arises with the RLS method when a constant forgetting factor is used, see e.g. (Åström and Wittenmark, 1995).

Different approaches to avoid windup include:

- Force the eigenvalues of $P(t)$ to lie in a given interval $[\lambda_{min}, \lambda_{max}]$. This can be achieved by making a factorization $P = UDU^T$ (using Bierman's UD factorization algorithm for example) and increasing D by dividing by the forgetting factor and limit the result to the specified interval. This algorithm is called selective forgetting (SF) and is described in (Parkum *et al.*, 1992). An algorithm with similar properties is proposed in (Cao and Schwartz, 1999).
- (Andersson and Broman, 1998) suggested to use a forgetting factor algorithm with different forgetting factors in the parameter recursion and P recursion. Faster forgetting is applied to P to avoid windup, which is motivated by speech signals.
- Add a positive definite matrix to the information matrix $P^{-1}(t)$, to assure that it is always invertible (so called Levenberg-Marquardt regularization). This method is discussed in (Gunnarsson, 1996).
- In the algorithm suggested by (Hägglund, 1983) the forgetting of old data is made in the same direction as incoming data and in such a way that the matrix P is driven toward a matrix proportional to the identity matrix when there is poor excitation.
- Consider the filter as a control system, where the goal is to achieve a pre-specified $P(t)$. One possibility is to use

$$Q(t) = \frac{P_d\varphi(t)\varphi^T(t)P_d}{R(t)+\varphi(t)^T P_d \varphi(t)}. \qquad (4)$$

  where P_d is the desired convergence point for the matrix P. This approach is in the sequel called the adaptive Kalman filter (AKF) algorithm.

2. THEORY

We will here take a closer look at the discrete time Ricatti equation (3d) in the adaptive algorithms. We have here restricted ourself to the case $n_y = 1$ to simplify notation. The idea with using (4) is to expand the 'covariance' ellipse only in the direction where excitation comes, and keeping the filter's sensitivity in the other directions and thus avoiding (eigenvalue) windup. Now P_d is easily seen to be a stationary point of the Ricatti equation, so

$$P_d = P_d - \frac{P_d\varphi(t)\varphi(t)^T P_d}{R(t)+\varphi(t)^T P_d\varphi(t)} + Q(t).$$

The matrix inversion lemma applied to $P_d - Q(t)$ gives

$$(P_d - Q(t))^{-1} = P_d^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T$$
$$\Rightarrow Q(t) = P_d - (P_d^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1}.$$

Similarly, the original recursive solution to the least squares equations can be written (or, apply the matrix inversion lemma in the same way as above):

$$(P(t+1) - Q(t))^{-1} = P(t)^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T$$
$$\Rightarrow Q(t) = P(t+1) - (P(t)^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1}.$$

By equating these two expressions for $Q(t)$, we get

$$\begin{aligned}
&P(t+1) - P_d = \\
&(P(t)^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1} \\
&\quad - (P_d^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1} \\
&= (P(t)^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1} \\
&\quad \times \left(P_d^{-1} - P(t)^{-1}\right)(P_d^{-1} + \varphi(t)R(t)^{-1}\varphi(t)^T)^{-1} \\
&= (I + P(t)\varphi(t)R(t)^{-1}\varphi(t)^T)^{-1}(P(t) - P_d) \\
&\quad \times (I + \varphi(t)R(t)^{-1}\varphi(t)^T P_d)^{-1}.
\end{aligned}$$

Denote $A(P) = I + P(t)\varphi(t)R(t)^{-1}\varphi(t)^T$. Then the covariance error $e(t) = P(t) - P_d$ evolves as

$$e(t+1) = A(P(t))^{-1}e(t)A(P_d)^{-T}. \qquad (5)$$

The stability of (5) can be concluded from Theorem 7.4 of (Jazwinski, 1970). From this theorem we know that the time-varying Kalman filter is uniformly exponentially stable provided the underlying system (2) is completely observable and controllable. This is true even if $Q(t)$ and $R(t)$ in (3) are not reflecting the true covariances since the result is algebraic in nature. Since the state transition matrix is the unit matrix the conditions are the following. There exists positive constants c, C and N such that

$$cI \le \sum_{j=k}^{k+N} \varphi(j)\varphi^T(j) \le CI, \quad \forall k \qquad (6)$$

$$cI \le \sum_{j=k}^{k+N} Q(j) \le CI, \quad \forall k \qquad (7)$$

The lower bound in (6) is the usual persistence of excitation condition. If (6) is satisfied and $P_d(0) > 0$ it follows from (4) that also (7) is satisfied.

The error recursion can be seen as a closed loop control system. The closed loop is approximately (since $A(P(t))$ depends on the controlled quantity) a first order system, which means that we can interpret the controller as a P-controller, controlling the system in Figure 1.

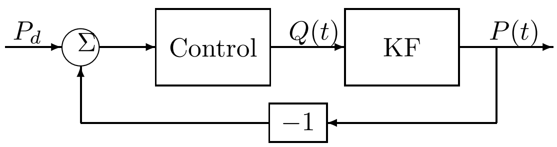

Fig. 1. Adaptive filtering as a control problem

3. NUMERICAL ILLUSTRATION

A simplified problem, yet common in practice, is to fit a straight line to two-dimensional data using the model:

$$y(t) = a\varphi_1(t) + b\varphi_2(t) + e(t). \tag{8}$$

As one of our motivations, (8) can model the control valves in a series of cascade coupled flotation tanks. Here $\varphi_1(t) = \sqrt{h(t)}$ and $\varphi_2(t) = u(t)\sqrt{h(t)}$ where $u(t)$ is the control signal for the valve and $h(t)$ is the level height difference over the valve (a measured quantity). The output $y(t)$ is the flow through the valve. This model has been found to give a good local description in the vicinity of some operating point. To adapt the model for different working points (which changes at unknown time instances) an algorithm for tracking of the time varying parameters a and b can be employed. Another use of parameter estimation is for detection of deviations in the parameters from the nominal ones, for which the controller was designed. The reason can e.g be wear or mechanical problems in the valve. Too large deviations would then indicate a severe loss in control performance. Here windup is certainly a problem. When the level set points in the tanks are constant and no or only small disturbances are acting on the process very little information is gained about the parameter values. It is then necessary to reduce the rate of forgetting of old data to avoid an increased noise sensitivity an consequently inaccurate estimates in certain directions. The consequences of this phenomenon would otherwise be a large bias (caused by noise in measured level h) and a stochastic uncertainty in the orientation of the plane (8) (we only know that it goes through the point $(y(t), \varphi_1(t), \varphi_2(t))$ but not the inclination).

An example of a realization is shown in Figure 2, where for simplicity $\varphi_1(t) \equiv 1$ and $\varphi_2(t) = u(t)$. The chosen $u(t)$ is poorly exciting around sample numbers 75 and 175, respectively. We will throughout this study assume that the true parameters are constant in time for simplicity, and comment on the implications of the results for the time-varying case. The off-line parameter estimates are illustrated in Figure 2 both as a linear model in a scatter plot of data and in the parameter space with a covariance ellipse. The covariance matrix $P(t)$ is only instrumental here, and should be interpreted as a time constant of the filter; in directions where it is thin, the time constant is long and where it is thick is is short, leading to a too sensitive estimator.

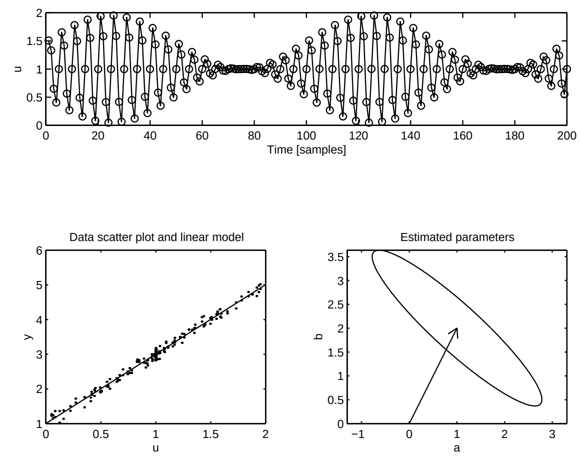

Fig. 2. Input $u(t)$ to the model (8), scatter plot of $(u(t), y(t))$ with an off-line estimated linear model, and estimate $\hat{\theta}_N$ in the parameter (a, b) space with uncertainty ellipse P_N.

Using the Kalman filter as a recursive estimator with $Q = 0.2I_2$ gives the result in Figure 3. The dashed line shows the 'confidence interval' of each parameter, which does not reflect the fact that the uncertainty of the difference of the parameters is poor. The last subplot shows the eigenvalues of $P(t)$, which better illuminate the lack of excitation around 75 and 175, respectively. A snapshot of the estimator at sample 30 and 75, respectively, shows that the former one has a more circular covariance shape, while the latter is uncertain of the difference $a - b$, see Figure 4. Note that the estimation error is clearly biased in this direction for this noise realization as can be expected to be the case in average.

4. DIFFERENT TECHNIQUES TO AVOID WINDUP

The goal of avoiding windup is to prevent an unbounded increase of the eigenvalues of the matrix P in case of poorly exciting signals. A lot of techniques have been suggested in the literature, many of them however very closely related. Here we compare the proposed AKF algorithm with

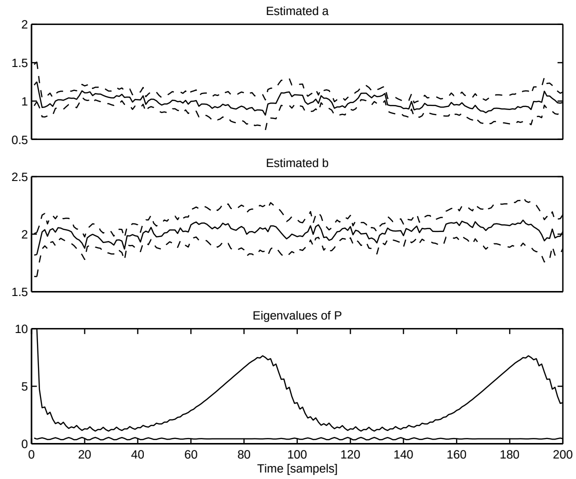

Fig. 3. Recursive parameter estimates with 'confidence intervals' as defined by $\sqrt{P^{(i,i)}(t)}$, and a plot of the eigenvalues of $P(t)$.

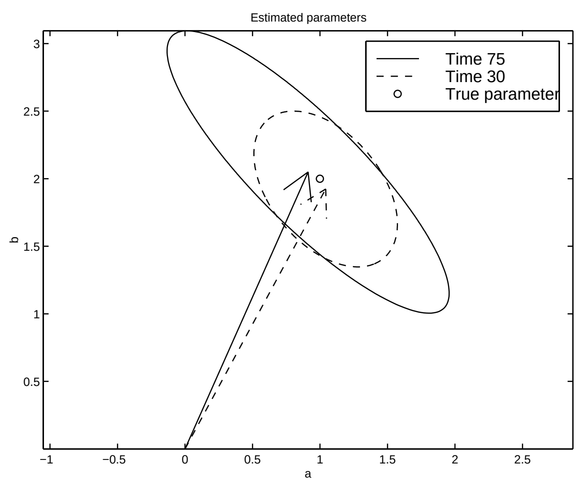

Fig. 4. Snapshot of parameter estimate at two time instants when KF is used.

the SF algorithm mentioned in the introduction. The design parameters are chosen to $\lambda_{min} = 0.8$, $\lambda_{max} = 1.2$ and $P_d = I_2$. These parameter choices have been made such that the 'confidence interval' is approximately the same for all the three algorithms when there is excitation.

Figures 5–8 show the resulting estimates. As compared to the Kalman filter algorithm it is obvious that in these cases the blow up of the 'covariance' matrix is prevented. The trade of between noise attenuation and tracking ability is depending on the eigenvalues plotted. This can be realized by studying the dynamics for the estimation error

$$\tilde{\theta}(t) = \theta^0(t) - \hat{\theta}(t)$$

where $\theta^0(t)$ is the true parameter vector at time t. From (3a)–(3d) we get

$$\tilde{\theta}(t) = (I - \bar{P}(t)\varphi(t)\varphi^T(t))\tilde{\theta}(t-1) + v(t-1) - \bar{P}(t)\varphi(t)e(t)$$

where $\bar{P}$ is given by

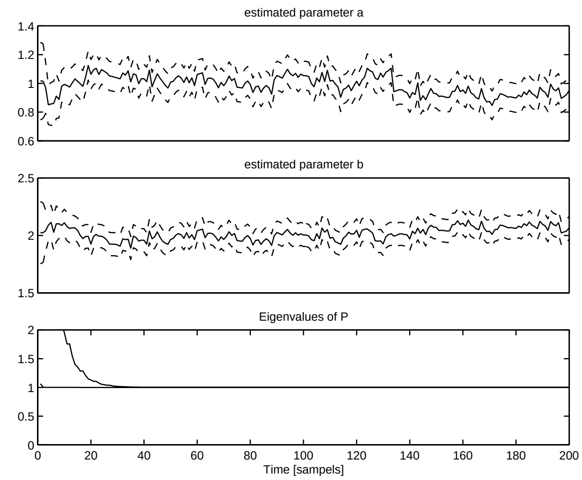

Fig. 5. Recursive parameter estimates using the AKF algorithm with 'confidence intervals' as defined by $\sqrt{P^{(i,i)}(t)}$, and a plot of the eigenvalues of $P(t)$. Here $P_d = I$.

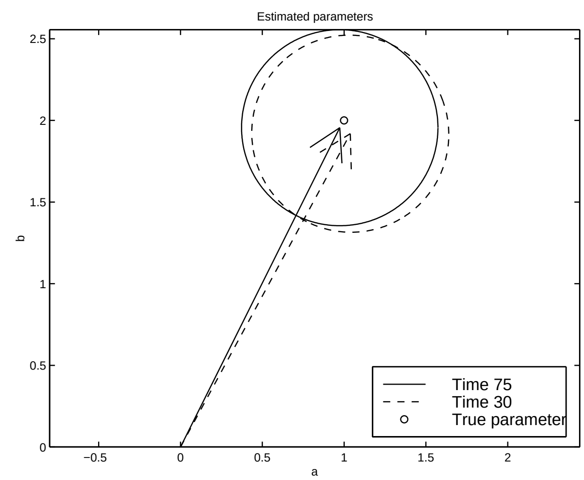

Fig. 6. Snapshot of parameter estimate at two time instants when the AKF algorithm is used.

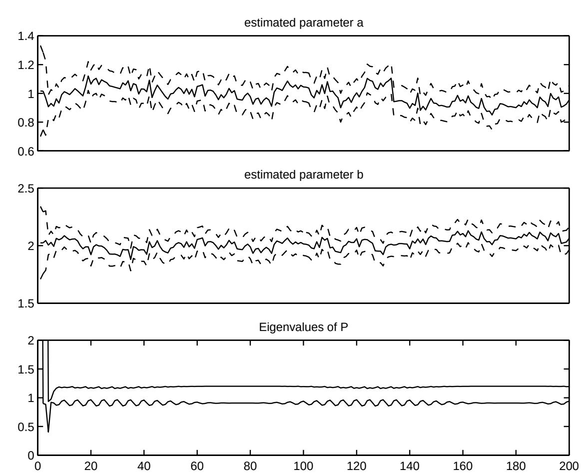

Fig. 7. Recursive parameter estimates using the SF algorithm with 'confidence intervals' as defined by $\sqrt{P^{(i,i)}(t)}$, and a plot of the eigenvalues of $P(t)$.

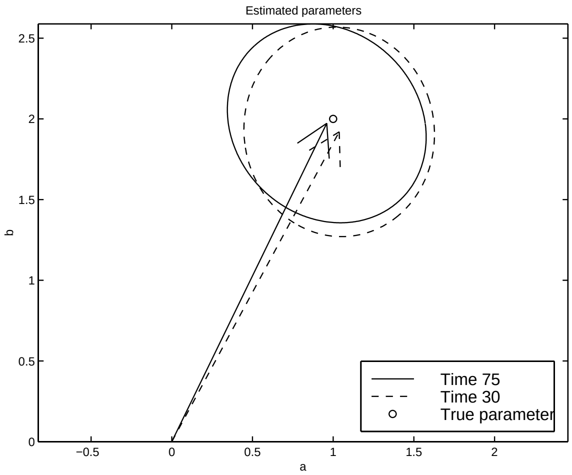

Fig. 8. Snapshot of parameter estimate at two time instants when the SF algorithm is used.

$$\bar{P}(t) = \frac{P(t-1)}{R(t) + \varphi(t)^T P(t-1)\varphi(t)}$$

Taking expectation of both sides neglecting certain dependencies we get

$$E\tilde{\theta}(t) = E(I - \bar{P}(t)Z)\tilde{\theta}(t-1)$$

where $Z = E\varphi(t)\varphi^T(t)$.

It is clear that for the KF algorithm the eigenvalues are inevitable fluctuating, the size depending on the excitation level, whereas the behaviour for SF and AKF algorithm is quite similar. The AKF algorithm shows however a smoother convergence of the eigenvalues.

5. TRACKING ABILITY

In this section the tracking behavior is studied. The b parameter in (8) is changed at two different time instants, one when there is good excitation at $t = 30s$ and the other when excitation is poor at $t = 75s$. The result is shown in figure 9.

The conclusions are that we automatically get better tracking when we have good excitation. Compare this to standard adaptive algorithms (RLS, LMS, KF) where the tracking time constant is basically fixed (for RLS this is completely true), and the uncertainty decreases when the excitation is good. These are two completely different paradigms.

6. THE BENEFIT OF CHANGE DETECTION

The compromise between noise attenuation and tracking speed is fundamental in adaptive filtering. If one wishes to have accurate estimates (i.e good noise attenuation) it is necessary to average over many data points which implies a low forgetting rate and consequently a slow filter. To speed

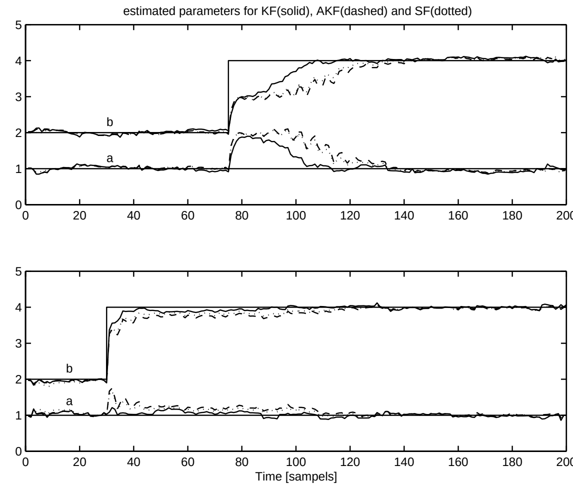

Fig. 9. Tracking ability of the KF (solid), AKF (dashed) and SF (dotted).

up the adaptation rate in case of fast changes in the true parameters it is therefore desirable to combine the filter with a change detector. When an abrupt change is detected the forgetting rate is temporarily increased and the estimates converge faster to the true values. Here the CUSUM detector, see e.g (Basseville and Nikiforov, 1993) and (Gustafsson, 2000), is employed, which is given by

$$g(t) = \max(g(t-1) + s(t) - v, 0)$$
$$\text{if } g(t) > h, \quad \text{then alarm and } g(t) = 0$$

where $g(t)$ is the *test statistics*, h is the *alarm threshold*, $s(t)$ is an input *distance measure* and ν is a *drift parameter* necessary to prevent false alarm due to the random walk drift in $g(t)$ when $s(t)$ is white noise. To detect abrupt parameter changes we here use $s(t) = e(t)$. This means that we will detect a change in the mean of the residuals in the filter.

One important question is to decide how to modify the tracking algorithm when an abrupt change has been detected. In this case when the change is known to occur in the b parameter (i.e knowledge based fault isolation) a natural modification is to increase $P(2,2)$ by a factor of 10 when a fault have been detected. This gives a fast fault identification.

In Figure 10 the tracking behaviour for the three algorithms are shown when combined with CUSUM detector (alarm times 76s and 31s respectively). As compared to Figure 9 the tracking speed is clearly increased when the jump occurs, most notable when excitation is poor.

7. CONCLUSIONS

Conventional adaptive algorithms for recursive parameter estimation, as recursive least squares, least mean square and the Kalman filter give a

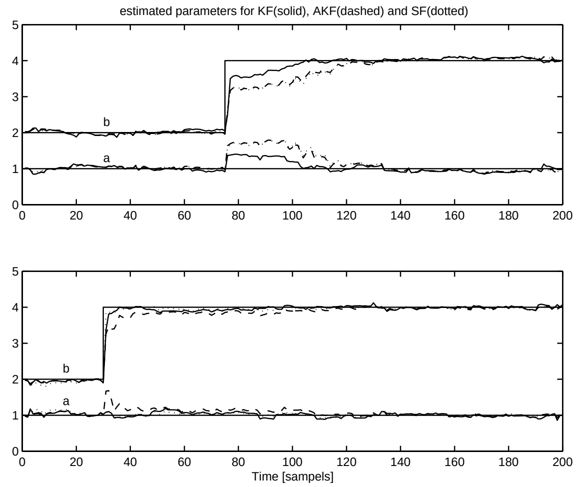

Fig. 10. Tracking ability of the KF (solid), AKF (dashed) and SF (dotted) algorithm when CUSUM detector is used.

filter with basically constant tracking speed and a parameter uncertainy that varies with excitation. The better excitation, the better estimates. In many applications the excitation varies a lot over time, and one wants to avoid drift in the estimates and preferably get a specified estimation accuracy in case the parameters are constant. This is achieved with the random walk covariance matrix in (4) and using a KF. The price to achieve this, is that the tracking gain varies with excitation, so it takes longer time to track changes when the excitation is poor.

However, the algorithm is computational simple (as compared to many other proposed algorithms), and intuitive to tune, since one decides the (relative) accuracy of the estimated parameters directly. Furthermore, this is exactly what is needed in change detection, where one wants to fix the false alarm rate (FAR), and then one takes the power to detect changes and time to detection the chosen FAR implies. The FAR is computed when the parameters do not change, and for that reason we want the parameter uncertainty to be constant and independent of excitation, and thus we accept longer time to detection and larger required changes when the excitation is poor. This again motivates the use of AKF.

The proposed concept was evaluated and compared to other approaches in a simulation study, and we pointed out the possibility to use change detection to speed up tracking.

8. REFERENCES

Anderson, B.D.O and J.B. Moore (1979). *Optimal Filtering*. Prentice Hall, Inc., Englewood Cliffs, N.J.

Andersson, Å. and H. Broman (1998). A second-order recursive algorithm with applications to adaptive filtering and subspace tracking. *IEEE Transactions on Signal Processing* **46**, 1720–1725.

Basseville, M. and I.V. Nikiforov (1993). *Detection of Abrubt Changes -Theory and Application*. PTR Prentice Hall.

Cao, L. and H.M. Schwartz (1999). A novel recursive algorithm for directional forgetting. In: *Proceedings of the American Control Conference*. pp. 1334–1338.

Gunnarsson, S. (1991). On some updating strategies in recursive identification. Report LiTH-ISY-I-1226. Linköping University.

Gunnarsson, S. (1996). Combining tracking and regularization in recursive least squares identification. In: *Proceedings of the 35th IEEE Conference on Decision and Control*. Kobe, Japan. pp. 2551–2552.

Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*. Wiley.

Hägglund, T. (1983). New Estimation Techniques for Adaptive Control. Doctoral dissertation. Department of Automatic Control, Lund Institute of Technology, Sweden.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering theory*. Academic Press, New York.

Ljung, L. and S. Gunnarsson (1990). Adaption and tracking in system identification - a survey. *Automatica* **26**, 7–21.

Parkum, J.E., N.K. Poulsen and J. Holst (1992). Recursive forgetting algorithms. *Int. Journal of Control* **55**, 109–128.

Åström, K.J. and B. Wittenmark (1995). *Adaptive Control, second edition*. Addison Wesley.