BAYESIAN ADAPTIVE FILTERING AT LINEAR COST

Tayeb Sadiki, Dirk T.M. Slock

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Tel: +33 4 9300 2656/2606; Fax: +33 4 9300 2627
e-mail:{sadiki,slock}@eurecom.fr

ABSTRACT

Standard adaptive filtering algorithms, including the popular LMS and RLS algorithms, possess only one parameter (step-size, forgetting factor) to adjust the tracking speed in a non-stationary environment. Furthermore, existing techniques for the automatic adjustment of this parameter are not totally satisfactory and are rarely used. In this paper we pursue the concept of Bayesian Adaptive Filtering (BAF) that we introduced earlier, based on modeling the optimal adptive filter coefficients as a stationary vector process, in particular a diagonal AR(1) model. Optimal adaptive filtering with such a state model becomes Kalman filtering. The AR(1) model parameters are determined with an adaptive version of the EM algorithm, which leads to linear prediction on reconstructed optimal filter correlations, and hence a meaningful approximation/estimation compromise. The resulting algorithm, of complexity $O(N^2)$, is shown by simulations to have performance close to that of the Kalman filter with true model parameters. In this paper, we apply a component-wise EM approach to further reduce the complexity to being linear in the number of adaptive filtering coefficients. The good performance of the resulting algorithm is illustrated in simulations. The AR(1) state model can be further approximated by a random walk, leading to further simplified adaptive filter that can be interpreted an LMS algorithm with a variable step-size per filter tap.

1. INTRODUCTION

In Bayesian Adaptive Filtering (BAF) [1], the evolution of filter coefficients is modeled as a stationary process. A simple choice for the search process is a first-order autoregressive process (AR(1)). This AR(1) model can be considered a state model. Hence, Bayesian Adaptive Filtering leads to Kalman filtering. This Kalman filtering needs to be adaptive because the model parameters are unkown. Even though adaptive Kalman filtering is a difficult problem, a surprisingly large number of solutions exist. The following approaches can be identified:

1. Recursive Prediction-Error Method (RPEM)
2. Extended Kalman Filter (EKF)
3. Best Quadratic Unbiased Estimator (BQUE)
4. Expectation-Maximization (EM)
5. Second-Order Statistics (SOS)

Eurécom's research is partially supported by its industrial partners: Hasler Foundation, Swisscom, Thales Communications, ST Microelectronics, CEGETEL, France Télécom, Bouygues Telecom, Hitachi Europe Ltd. and Texas Instruments. The work reported herein was also partially supported by a PACA regional scholarship.

6. Subspace-Based Estimation Method (SBEM)

A common approach is the well-known Recursive Prediction-Error Method (RPEM), which provides an estimator that minimizes a prediction error criterion function $V_N(\theta)$, of the form

$$\hat{\theta} = arg \min_{\theta} V_N(\theta) \tag{1}$$

where θ is the set of parameters to be estimated. However, for many scenarios (1) has no closed-form solution, due to non convexity of $V_N(\Theta)$ in Θ. A popular choice for solving the optimization problem are gradient-based search techniques, and therefore implementation complexity becomes similar to the ML approach. A standard state estimation method used for polynomial systems is the Extended Kalman Filter (EKF), which allows simultaneous estimation of states and parameters through a Recursive prediction-correction model [2]. As an approximate conditional mean filter, the EKF is suboptimal. A popular and robust alternative to these algorithms is provided by the subspace-based estimation methods [3]. These algorithms extract the estimates of system state-space matrices directly from data by first dividing that data into past and future data and then projecting the future data onto the space spanned by the past data. A bank of Kalman filters is employed to compute the estimation of the state sequence, which results in an approximation of Kalman filter estimate of the state. A good alternative to the described schemes is given by the EM algorithm, where the estimate of the state sequence is found by a single Kalman smoothed estimation instead. In this case, the smoothed state estimates are calculated under the assumption that the parameters of the true system are the same as the current estimate. Other approaches like Second-Order Statistics (SOS) methods and Best Quadratic Unbiased Estimator (BQUE) can be found in [4]. In our work we focus on EM parameter estimation techniques.

Previous works on suboptimal adaptive filtering coefficient estimation [5],have shown how an RLS (Recursive Least Squares) approach can successfully converge at fast speed. However, this approach is rather complex, presenting computation complexity of order $O(N^2)$. A less compless solution can be found by using a LMS (Least Mean Squares) [6] , [7], [8], which reduces the complexity of RLS [9] algorithms in one order of magnitude, resulting $O(N)$. On the other hand, while complexity is reduced convergence speed of LMS is slower than the one shown by RLS algorithms. In these techniques, the optimal adaptive filter is considered to be an unknown deterministic function, which in reality is not true due to channel randomness. Since the introduction of the LMS algorithm by Widrow and Hopf in the 1960's, most

of the further work in adaptive filtering has focused on improving theinitial convergence. The Recursive Least-Squares (RLS) algorithm was also developed in the 1960'sand provided an alternative algorithm for adaptive system identification. The RLS algorithm is recursive and not iterative as the LMS algorithm, solving a LS cost function exactly at each update. As a result it converges very fast since it provides an unbiased solution once the LS problem gets overdetermined. This deterministic aspect adds up to the observation that the RLS convergence is insensitive to the input signal correlation structure (approximately, since there is some dependence on the initialization). The RLS algorithm, though providing computational savings w.r.t. the plain solving of LS problems at each sampling period, is quite a bit more expensive than the LMS algorithm. The KF'ing framework can be straightforwardly extended to incorporate time-varying optimal parameters. The simplest way is probably through the following $AR(1)$ model state equation for optimal filter variation

$$y_k = X_k^H H_{k-1}^0 + v_k \quad (2)$$

where X_k is a $N \times 1$ input vector of white complex symbols and N is the length of the filter. The noise v_k is assumed to be zero mean uncorrelated and normally distributed with common covariance matrix R. The Bayesian Filter H_k^0 is assumed to be of primary interest [1]; it is modeled as a first order multivariate process of the form

$$H_k^0 = AH_{k-1}^0 + W_k \quad (3)$$

where $E[W_k W_k^H] = Q_k$ is a $N \times N$ transition matrix describing the way the underlying series move through successive time periods. The BF H_k^0 may be non-stationary since we do not make special assumptions about the roots of the characteristic equation A. The $N \times 1$ noise terms W_k, are zero-mean uncorrelated normal vectors with common covariance matrix Q.

The motivation for the model defined by (2) and (3) originates from a desire to account separately for uncertainties in the model as defined by model error W_k and uncertainties in measurements made on the model as expressed by the measurement noise process v_k. It might be helpful to envision (2) as a kind of random-effects model for the time variation, where the effect vector H_k^0 has a correlation structure over time imposed by the multivariate autoregressive model (3). In this context, it is a generalization of the ordinary autoregressive AR model which accounts for observation noise as well as model induced noise.

In this paper, we provide a convenient method for dealing with the incomplete data problem. To estimate H_k^0 we use Kalman filtering with one-step smoothing. The primary aim of a smoothing procedure is to estimate the unobserved time-varying H_k^0. If one knows the values for the parameters Q and A the conventional Kalman smoothing estimators can be calculated as conditional expectations and will have MMSE .

Since the smoothed values in a Kalman filter estimator will depend on the initial values assumed for the above parameters, it is of interest to consider various ways in which they might be estimated. In most cases this has been accomplished by Maximum likelihood techniques involving the use of scoring or Newton-Raphson technique to solve the nonlinear equations wich result from differentiating the log-likelihood function. In this paper, we introduce an EM approach for iteratively update the parameter model. Experimental results will be shown for the proposed algorithm, comparing to KF filtering.

2. PARAMETER ESTIMATION VIA THE EM ALGORITHM

In this section we develop the EM algorithm for estimating the parameters of (2)- (3). Perhaps the most important step in applying the EM algorithm to a particular problem is that of choosing the missing data. The missing data should be chosen so that the task of maximizing $U(\theta, \theta^{(k)})$ for any value of $\theta_k = (A, Q)$ is easy and so that it is possible to perform the expectation step.

Fortunately, in this case, the choice of missing data is not too difficult. Let us imagine for a moment that, in addition to the system inputs and outputs, X_k and Y_k respectively, the state H_k^0 was available. In that case, ML estimation of A reduces to $arg\min_A ||H_k^0 - AH_{k-1}^0||_Q^2$. The covariance elements of W_k, Q, could then be calculated from the residuals. Moreover, the conditional expectation of state sequence may be calculated using a (slightly augmented) Kalman Smoother. All of this suggests that the state sequence is a desirable conditionate for the missing data. We therefore designate Y as the incomplete data so that the complete data set is $Z = (H_k^0, Y_k)$.

In order to develop a procedure for estimating the parameters in the state-space model defined by (2)- (3), we note first that applying the Bayes rule, the probability density associated with Z can be written in the form

$$f_Z(z, \theta) = f_{Z|Y=y}(z, \theta) . f_Y(y; \theta) \quad (4)$$

where $f_Y(y; \theta)$ is the probability density function of Y and $f_{Z|Y=y}(z, \theta)$ is the conditional probability density of Z given $Y = y$. Taking the logarithm on both sides of (4), we get the log-likelihood function

$$\log f_Y(y, \theta) = \log f_Z(z, \theta) - \log f_{Z|Y=y}(z, \theta) \quad (5)$$

Define, for convenience

$$L(\theta) = \log f_Y(y, \theta)$$

With this definition we can write

$$\begin{aligned} L = & -2\log f_\theta(H_k, Y_M, \theta | Y_M) \\ & + M \log det Q \\ & + \sum_{k=1}^{M} tr(H_k - AH_{k-1}) Q^{-1} (H_k - AH_{k-1})^H \\ & + \sum_{k=1}^{M} tr(y_k - X_k^H H_{k-1}) R^{-1} (y_k - X_k^H H_{k-1})^H \end{aligned} \quad (6)$$

The log-likelihood given above depends on the unobserved data H_k^0. We consider applying the EM algorithm conditionally with respect to the observed ensemble Y. That is, the estimated parameters at the $(k+1) - th$ iteration are the values A and Q that maximize

$$U(\theta, \hat{\theta}_k) = E_{\hat{\theta}_k} \{\log f_\theta(H_k, Y_M, \theta | Y_M)\} \quad (7)$$

where $E_{\hat{\theta}_k}$ denotes the conditional expectaion relative to a density containing the $k-th$ iteration values.
In order to calculate the conditional expectation defined in (**??**), it is convenient to define the conditional mean

$$\hat{H}_k = E_{\hat{\theta}_k}\{H_k^0|Y_M\}$$
$$P_k = E[\tilde{H}_k\tilde{H}_k^H|Y_M\}$$
$$P_{k-1} = E[\tilde{H}_{k-1}\tilde{H}_{k-1}^H|Y_M\}$$

we suppose the following definitions

$$\begin{aligned}
\mathbf{\Pi_{k|k}} &= \sum_{k=1}^{M}(E_{\hat{\theta}_k}\{H_{k-1}^0(H_{k-1}^0)^H|Y_M\}+P_{k-1}) \\
\mathbf{\Pi_{k-1|k}} &= \sum_{k=1}^{M}(E_{\hat{\theta}_k}\{H_k^0(H_k^0)^H|Y_M\}+P_k) \\
\mathbf{\Pi_{k,k-1|k}} &= \sum_{k=1}^{M}(E_{\hat{\theta}_k}\{H_k^0(H_{k-1}^0)^H|Y_M\}+P_{k,k-1}) \quad (8)
\end{aligned}$$

The Kalman filter terms $\hat{H}_k$, P_k and $P_{k,k-1}$ are computed under the parameter values A_k and Q_k using the recursions in (8). Maximizing (7) w.r.t. A and Q, we obtain

$$\begin{aligned}
\mathbf{Q_{k+1}} &= \frac{1}{M}(\mathbf{\Pi_{k|k}} - \mathbf{\Pi_{k,k-1|k}}(\mathbf{\Pi_{k-1|k}})^{-1}\mathbf{\Pi_{k,k-1|k}^H}) \\
\mathbf{A_{k+1}} &= \mathbf{\Pi_{k,k-1|k}}(\mathbf{\Pi_{k-1|k}})^{-1}
\end{aligned}$$

3. ADAPTIVE EM-KALMAN ALGORITHM

In our study, the tasks of smoothing in a missing data context are interpreted as basically the problem of estimating the BAF H_k^0 in the state-space model (2)- (3). The conditional means provide a minimum MSE solution based on the observed data. The parameters Q and A are estimated by ML using the EM algorithm. We simplify the estimation problem by considering A and Q diagonal matrices. The filter parameters are iteratively computed through M iterations. The estimation of the optimal filter variation is carried out by KF'ing and one step smoothing and we introduce an EM approach to iteratively update the parameter model.
The resulting algorithm is shown in Table of Adaptive EM-Kalman filter. The complexity of Kalman filter is limited to $O(N^2)$ order and the Adaptive Kalman filter has the same order of complexity. To reduce the complexity of our algorithm we propose two methods. First, a cyclic minimization method by using a Maximum A posteriory and Maximum Likelihood estimation (MAP-ML), MAP for estimating the Bayesian filter sequence and ML for estimating the parameters model. Second, a Component-Wise Adaptive Kalman filter, which is based on the estimation of each parameter one by one.

4. MAP-ML ESTIMATION

The value of H_k^0 that maximizes the posterior density (that is, the mode of the posterior density) is called the maximum a posterior probability estimate of H_k^0.
If the posterior density of H_k^0 given A, Q and Y is unimodal and symmetric, then it is easy to see that the MAP estimate and the mean squared estimate coincide, since the posterior density attains its maximum value at its expected value.
Let the sequence filter H_k^0 be considered as a random variable distributed according to the posterior density $f_{H_k^0}(h_k^o)$. The posterior distribution for H, is given by

$$f_{H_k^0,Y|A,Q}(h_k^o, y \mid A, Q) \quad (9)$$

then $\widehat{H_{kMAP}}$ is obtained by maximizing the logarithm of the posterior density with respect to H_k^0. Initially, A_0 and Q_0 are set to a certain initial value. After the first iteration, A_{k+1} and Q_{k+1} are obtained by ML, given $\widehat{H}_{k,MAP}$.

5. COMPONENT-WISE ADAPTIVE KALMAN ALGORITHM

Our goal is to design an optimal algorithm with reduced complexity in a realistic environment, considering the filter coefficients to estimate random variables. In a previous section, a Bayesian Adaptive Filtering (BAF) approach has been proposed, showing a complexity of order $O(N^2)$. To reduce the complexity of the algorithm presented in Table of Adaptive EM-Kalman filter, we propose a Component-Wise Adaptive Kalman algorithm to update the filter coefficients, which decreases computational complexity in 1 order of magnitude while preserving convergence. Experimental results will be shown for the proposed algorithm, comparing to KF filtering and Adaptive Kalman algorithms.The filter parameters are iteratively computed through M iterations. The system (2)-(3) becomes for $n = 1 \ldots N$, where N is the length of the filter

$$\begin{aligned}
h_{k,n}^o &= a_n h_{k-1,n}^o + w_{k,n} \quad (10) \\
y_k &= h_{k-1,n}^o x_{k,n} + \sum_{j\neq n}^{N} h_{k-1,n}^o x_{k,n} + v_k \quad (11)
\end{aligned}$$

and

$$h_k^o = \hat{h}_k + \tilde{h}_k$$

we can write

$$y_k - \sum_{j\neq n}^{N}\hat{h}_{k-1,n}^o x_{k,n} = h_{k-1,n}^o x_{k,n} + \sum_{j\neq n}^{N}\tilde{h}_{k-1,n}^o x_{k,n} + v_k$$

In each iteration y_k and v_k are updated as follows

$$y_k' = y_k - \sum_{j\neq n}^{N}\hat{h}_{k-1,n}^o x_{k,n}$$

and

$$v_k' = \sum_{j\neq n}^{N}\tilde{h}_{k-1,n}^o x_{k,n} + v_k$$

The missing data should be chosen so that the task of maximizing $U(\theta, \theta_n^k)$ for $n = 1 \ldots N$ $\theta_n^l = (a_n, q_n)$ is easy and so that it is possible to perform the expectation step.
Fortunately, in this case, the choice of missing data is not too

difficult. Let us imagine for a moment that, in addition to the system inputs and outputs, $x_{k,n}$ and Y_k respectively, the state h_n^0 was avaible then ML estimation of a_n reduces to applying to (11). The covariance elements, q_n, of w_k could then be calculated from the residuals. Moreover, the conditional expectation of state sequence may be calculated using a (slightly augmented) Kalman Smoother. All of this suggests that the state sequence is a desirable conditionate for the missing data. We therefore designate Y as the incomplete data so that the complete data set is $Z = (h_n^o, Y)$.
For the $n-th$ iteration the log-likelihood function can be writen as

$$\begin{aligned} L &= -2\log f_\theta(h_n^o, Y_M, \theta|Y) \\ &= N\log det q_n \\ &\quad + \sum_{k=1}^{M} q_n^{-2}(h_{k,n}^o - a_n h_{k-1,n}^o)(h_{k,n}^o - a_n h_{k-1,n}^o)^H \\ &\quad + \sum_{k=1}^{M} \sigma_v^{-2}(y_k - x_{k,n}^H h_{k-1,n}^o)(y_k - x_k^H h_{k-1,n}^o)^H \quad (12) \end{aligned}$$

where L is to be maximized with respect to parameters a_n and q_n . Since the log-likelihood given above depends on the unobserved data $h_{k,n}^o$, we consider applying the EM algorithm conditionally with respect to the observed Y. That is, the estimated parameters at the $(k+1)-th$ iterate as the values a_n and q_n which maximize

$$U(\theta, \hat{\theta}_k) = E_{\hat{\theta}_k}\{\log f_\theta(h_n^o, Y_M, \theta|Y)\} \quad (13)$$

where $E_{\hat{\theta}_k}$ denotes the conditional expectaion relative to a density containing the kth iterate values.
In order to calculate the conditional expectation defined in (13), it is convenient to define the conditional mean

$$\hat{h}_{k,n} = E_{\hat{\theta}_k}\{h_{k,n}|Y\}$$

and

$$P_{k,n} = E\{\tilde{h}_{k,n}\tilde{h}_{k,n}^H\}$$

we suppose the following definitions

$$\begin{aligned} \pi_{k,n|k} &= \sum_{k=1}^{M} E_{\hat{\theta}_k}\{h_{k-1,n}^o h_{k-1,n}^{oH}|Y\} + P_{k-1,n} \\ \pi_{k-1,n|k} &= \sum_{k=1}^{M} E_{\hat{\theta}_k}\{h_{k,n}^o h_{k,n}^{oH}|Y\} + P_{k,n} \\ \pi_{k,k-1,n|k} &= \sum_{k=1}^{M} E_{\hat{\theta}_k}\{h_{k,n}^o h_{k-1,n}^{oH}|Y\} + P_{k,k-1} \quad (14) \end{aligned}$$

The Kalman filter terms $\hat{h}_{k,n}$, $P_{k,n}$ and $P_{k,k-1}$ are computed under the parameter values $a_{n,k}$ and $q_{n,k}$ using the recursions in (12).Furthermore, it is easy to see that the choices

$$\begin{aligned} q_{k+1,n+1} &= \frac{1}{\gamma_k}(\pi_{\mathbf{k|k}} - \pi_{\mathbf{k,k-1|k}}(\pi_{\mathbf{k-1|k}})^{-1}(\pi_{\mathbf{k,k-1|k}})^{\mathbf{H}}) \\ a_{k+1,n+1} &= \pi_{\mathbf{k,k-1|k}}(\pi_{\mathbf{k-1|k}})^{-1} \end{aligned}$$

maximize the last two lines in the Expectation-likelihood function (13). In our study, the tasks of smoothing in a missing data context are interpreted as basically the problem of estimating the BAF $h_{k,n}^0$ in the state-space model (3)-(4). The conditional means provide a minimum MSE solution based on the observed data. The parameters q_n and a_n are estimated by ML using the component-wise EM algorithm. We simplify the estimation problem by considering a_n and q_n diagonal matrices. The filter parameters are iteratively computed through M iterations. The estimation of the optimal filter variation is carried out by KF'ing and one step smoothing and we introduce an EM approach for iteratively update the parameter model.
The algorithm is resulting in Table of component-wise Adaptive EM-Kalman.

6. NUMERICAL RESULTS

The behavior of Kalman and Adaptive Kalman filters are compared on the basis of simulation results, as shown in Fig. 1. The proposed algorithms are implemented with the initial parameters $\alpha = 0.9$ and forgeting factor $\lambda = 0.95$. The concept of Bayesian Adaptive Filtering (BAF) that we introduced earlier, based on modeling the optimal adaptive filter coefficients as a stationary vector process, is in particular a diagonal AR(1) model $H_k^0 = AH_{k-1}^0 + W_k$. Optimal adaptive filtering with such a state model becomes Kalman filtering. The AR(1) model parameters are determined with an adaptive version of the EM algorithm, which leads to linear prediction on reconstructed optimal filter correlations, and hence a meaningful approximation/estimation compromise. The optimal parameters are $A = 0.96 * I$,where I is identity matrix, and the error covaraince matrix Q is an exponential power delay profile, with the characteristic parameter $\beta = 0.9$. The input signal X is considered to be white.

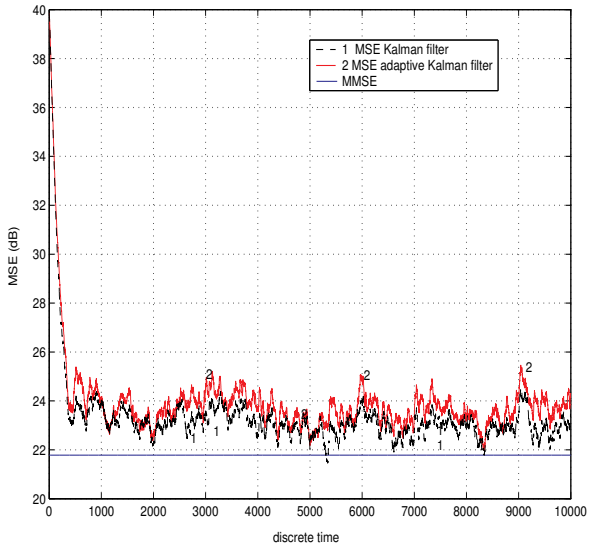

Figure 1: Comparison between the proposed Adaptive Kalman algorithm and Kalman filter

The behavior of CW-EM Adaptive Kalman, Adaptive Kalman and Kalman filter algorithms are compared on the basis of simulation results, as shown in Fig. 2

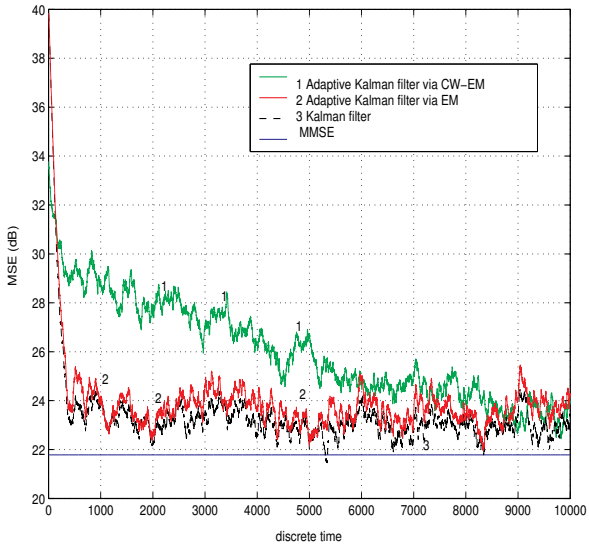

Figure 2: Comparison between the proposed CW-EM Adaptive Kalman algorithm and Adaptive Kalman and Kalman filter typical algorithms

7. SIMPLIFIED COMPONENT-WISE ADAPTIVE KALMAN ALGORITHM

We consider A to be an indentity matrix. Hence the comlexity of the adaptive part is comparable to the one exhibited by tap Variable Step-Size (TVSS) LMS [10], [6], [7], [8] like $4N$. In practice A tends to the identity matrix when MSE converges to $MMSE$. The process is low-pass which is equivalent to a random walk.

8. CONCLUSION

As Fig. 1 shows, the proposed Adaptive Kalman algorithm converges to the ML estimator. The convergence speed of the proposed algorithm in a random time-varying environment is approximately as fast as the one shown by conventional deterministic Kalman filtering (known parameters). In the proposed scheme, parameter estimation is carried out through the EM algorithm, hence assuring convergence to the ML estimator when a favorable initialization is provided.On the other hand, to take A and Q a digonal matrix, the complexity of Kalman filter is limited to $O(N^2)$ order and the Adaptive Kalman filtering have the some order of complexity. As Fig. 2 shows, the proposed Component-Wise adaptive Kalman algorithm converges to the ML estimator. The convergence speed is slower than in the cases where conventional Kalman adaptive filter or Kalman filter algorithm is applied. However, the proposed algorithm clearly outperforms the other approaches in terms of complexity.Hence, the complexity of Component-Wise adaptive Kalman filter is linear in N, the adaptive filter order. Never the less, upon convergence, the perfermance of the Component-Wise adaptive Kalman filter is comparable of the kown Kalman filter.

Adaptive EM-Kalman Algorithm	
Computation	**Cost ($\times$)**
Initialization	
$\hat{\mathbf{H}}_{\mathbf{0|0}}=\hat{\mathbf{0}},\ \mathbf{P}_{\mathbf{0|0}}=\mathbf{100I}$, $\mathbf{A_0}=\alpha\mathbf{I},\ \mathbf{Q_0}=(\mathbf{1}-\alpha)\mathbf{I}$ $\mathbf{\Pi_{0|0}}=\mathbf{0},\ \mathbf{\Pi_{1,0|0}}=\mathbf{0},\ \mathbf{\Pi_{1|0}}=\mathbf{0}$ $\gamma^{(0)}=0$	$2N$
Kalman filtering and one step smoothing	
$\hat{\mathbf{H}}_{\mathbf{k|k-1}}=\hat{\mathbf{A}}_{\mathbf{k}}\hat{\mathbf{H}}_{\mathbf{k-1|k-1}}$	N^2
$\hat{y}_{k\|k-1}=\mathbf{x}_{\mathbf{k}}^{\mathbf{H}}\hat{\mathbf{H}}_{\mathbf{k|k-1}}$	N^2
$\mathbf{K}_{\mathbf{k}}=\mathbf{P}_{\mathbf{k|k-1}}\mathbf{x}_{\mathbf{k}}$	N^2
$\mathbf{M}_{\mathbf{k}}=(\mathbf{x}_{\mathbf{k}}^{\mathbf{H}}\mathbf{K}_{\mathbf{k}}+\sigma_{\mathbf{v}}^{\mathbf{2}})^{-1}$	N
$\mathbf{K}_{\mathbf{k}}^{\mathbf{f}}=\mathbf{K}_{\mathbf{k}}\mathbf{M}_{\mathbf{k}}$	1
$\mathbf{C}_{\mathbf{k-1}}=\mathbf{P}_{\mathbf{k-1|k-1}}\mathbf{A}_{\mathbf{k}}^{\mathbf{H}}\mathbf{P}_{\mathbf{k|k-1}}^{-\mathbf{1}}$	$2N^2$
$\hat{\mathbf{H}}_{\mathbf{k-1|k}}=\hat{\mathbf{H}}_{\mathbf{k|k-1}}+\mathbf{K}_{\mathbf{k}}^{\mathbf{f}}(\mathbf{y_k}-\hat{\mathbf{y}}_{\mathbf{k|k-1}})$	1
$\mathbf{P}_{\mathbf{k|k-1}}=\mathbf{A_k}\mathbf{P}_{\mathbf{k-1|k-1}}\mathbf{A}_{\mathbf{k}}^{\mathbf{H}}+\mathbf{Q_k}$	$N(\frac{N-1}{2})$
$\hat{\mathbf{H}}_{\mathbf{k|k}}=\hat{\mathbf{H}}_{\mathbf{k|k-1}}+\mathbf{A}_{\mathbf{k}}^{-\mathbf{1}}(\mathbf{K_k}-\mathbf{Q_k}\mathbf{x_k})\mathbf{M_k}(\mathbf{y_k}-\hat{\mathbf{y}}_{\mathbf{k|k-1}})$	$N+1$
$\mathbf{P}_{\mathbf{k|k}}=\mathbf{P}_{\mathbf{k|k-1}}-\mathbf{K}_{\mathbf{k}}^{\mathbf{f}}\mathbf{K}_{\mathbf{k}}^{\mathbf{H}}$	1
$\mathbf{P}_{\mathbf{k-1|k}}=\mathbf{P}_{\mathbf{k-1|k-1}}+\mathbf{C}_{\mathbf{k-1}}(\mathbf{P}_{\mathbf{k|k}}-\mathbf{P}_{\mathbf{k|k-1}})$	N^2
Model Parameters Adaptation	
$\mathbf{\Pi_{k|k}}=\lambda\mathbf{\Pi_{k|k-1}}+\mathbf{diag}(\hat{\mathbf{H}}_{\mathbf{k|k}}\hat{\mathbf{H}}_{\mathbf{k|k}}^{\mathbf{H}}+\mathbf{P_{k|k}})$	N
$\mathbf{\Pi_{k-1|k}}=\lambda\mathbf{\Pi_{k-1|k-1}}+\mathbf{diag}(\hat{\mathbf{H}}_{\mathbf{k-1|k}}\hat{\mathbf{H}}_{\mathbf{k-1|k}}^{\mathbf{H}}+\mathbf{P_{k-1|k}})$	N
$\mathbf{D_k}=\mathbf{P_{k|k}}\mathbf{C}_{\mathbf{k-1}}^{\mathbf{H}}=\mathbf{A_k}\mathbf{P_{k-1|k-1}}-\mathbf{K}_{\mathbf{k}}^{\mathbf{f}}(\mathbf{A}_{\mathbf{k}}^{-\mathbf{1}}(\mathbf{K_k}-\mathbf{Q_k}\mathbf{x_k}))^{\mathbf{H}}$	$2N$
$\mathbf{\Pi_{k,k-1|k}}=\lambda\mathbf{\Pi_{k,k-1|k-1}}+\mathbf{diag}(\hat{\mathbf{H}}_{\mathbf{k|k}}\hat{\mathbf{H}}_{\mathbf{k-1|k}}^{\mathbf{H}}+\mathbf{D_k})$	N
$\mathbf{Q_{k+1}}=\frac{\mathbf{1}}{\gamma_{\mathbf{k}}}(\mathbf{\Pi_{k|k}}-\mathbf{\Pi_{k,k-1|k}}(\mathbf{\Pi_{k-1|k}})^{-\mathbf{1}}(\mathbf{\Pi_{k,k-1|k}})^{\mathbf{H}})$	$2N$
$\gamma_k=\gamma_{k-1}+1$	
$\mathbf{A_{k+1}}=\mathbf{\Pi_{k,k-1|k}}(\mathbf{\Pi_{k-1|k}})^{-\mathbf{1}}$	N
cost/update $7.5N^2+11.5N+3$	

Talbe.1: Adaptive EM-Kalman Algorithm

REFERENCES

[1] T. Sadiki and D. T. M. Slock, "Bayesian adaptive feltering: Principles and practical approaches," in *Eusipco Coference*, 16-17 September 2004.

[2] Torsten Soderstrom, *System Identification*, Prentice Hall Sgnal Prossing Series, 1989.

[3] B. Jansson, M.; Goransson and B. Ottersten, *A subspace method for direction of arrival estimation of uncorrelated emitter signals*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 47 , Issue: 4,945 - 956, April 1999.

[4] Jerry M. Mendel, *Lessons In Estimation Thory For Signal Prossing, Communications, And Control*, Prentice Hall Signal Prossing Series, 1995.

[5] Lars Lindbom, Mikael Sternad, and Anders Ahlen, "Adaptation with constant gains: Analysis for fast variations," in *Acoustics, Speech, and Signal Proce. ICASSP, IEEE International Conf. on , Vol. 2*, 2002.

[6] G. L. Cassio and M. B. Jose Carlos, "Evaluation and design of variable step size adaptive algorithms," in *Acoustics, Speech, and Signal Proce.ICASSP, IEEE International Conf. Vol.: 6*, May 2001.

[7] T. Aboulnaser and K. Mayyas, "A robust variable step-size lms-type algorithm: Analysis and simulations," in *IEEE Trans. on Signal Proces.vol. 45, no. 3*, March 1997.

[8] S. B. Gelfand, Y. Wei, and V. K James, "The stability of variable step-size lms algorithms," in *IEEE Trans. on Signal Proces.vol. 47, no. 12*, December 1999.

[9] L. Ljung L. Guo, "Performance analysis of the forgetting factor rls algorithm," in *Int. J. Adaptive Cont. Sig. Proces., vol.7, pp. 141-537*, 1993.

[10] W. Liu, "Performance of joint data and channel estimation using tap variable step-size (TVSS) lms for multipath fast fading channel," in *IEEE Trans. on Signal Proces.*, 1994.

Adaptive Component-Wise EM-Kalman Algorithm	
Computation	**Cost ($\times$)**
Initialization	
$\hat{h}_{0\|0}=0\,,\ P_{0\|0}=100\,,$ $a_0=\alpha\,,\ q_0=(1-\alpha)$ $\pi_{0\|0}=0\,,\ \pi_{1,0\|0}=0,\ \pi_{1\|0}=0$ $\gamma_0=0$	
Kalman filtering and one step smoothing	
$for\ n=1\ldots N$	
$\hat{h}_{k,n\|k-1}=a_{n,k}\hat{h}_{k-1,n\|k-1}$	1
$\hat{y}_{k,n\|k-1}=x^H_{n,k}\hat{h}_{k,n\|k-1}$	1
$K_{n,k}=P_{k,n\|k-1}x_k$	1
$M_{k,n}=(x^H_{k,n}K_{n,k}+\sigma^2_v)^{-1}$	1
$K^f_{n,k}=K_{n,k}M_{n,k}$	1
$C_{n,k-1}=P_{n,k-1\|k-1}a^H_{n,k}P^{-1}_{n,k\|k-1}$	2
$\hat{h}_{n,k-1\|k}=\hat{h}_{n,k\|k-1}+K^f_{n,k}(y_k-\hat{y}_{k,n\|k-1})$	1
$P_{k,n\|k-1}=a_{n,k}P_{k-1,n\|k-1}a^H_{n,k}+q_{n,k}$	1
$\hat{h}_{k,n\|k}=\hat{h}_{k,n\|k-1}+a_{n,k}(K_k-q_{n,k}x_k)M_k(y_k-\hat{y}_{k,n\|k-1})$	2
$P_{k,n\|k}=P_{k,n\|k-1}-K^f_{n,k}K^H_{n,k}$	1
$P_{k-1,n\|k}=P_{k-1,n\|k-1}+C_{n,k-1}(P_{n,k\|k}-P_{n,k\|k-1})$	1
Model Parameters Adaptation	
$\pi_{n,k\|k}=\lambda\pi_{k,n\|k-1}+diag(\hat{h}_{n,k\|k}\hat{h}^H_{n,k\|k}+P_{n,k\|k})$	1
$\Pi_{n,k-1\|k}=\lambda\pi_{n,k-1\|k-1}+diag(\hat{h}_{n,k-1\|k}\hat{h}^H_{n,k-1\|k}+P_{n,k-1\|k})$	1
$D_{n,k}=P_{n,k\|k}C^H_{n,k-1}=a_{n,k}P_{n,k-1\|k-1}-K^f_{n,k}(a^{-1}_{n,k}(K_{n,k}-q_{n,k}x_{n,k}))^H$	2
$\pi_{n,k,k-1\|k}=\lambda\pi_{n,k,k-1\|k-1}+diag(\hat{h}_{n,k\|k}\hat{H}^H_{n,k-1\|k}+D_{n,k})$	1
$q_{n,k+1}=\frac{1}{\gamma_k}(\pi_{n,k,\|k}-\pi_{n,k,k-1\|k}(\pi_{n,k-1\|k})^{-1}(\pi_{n,k,k-1\|k})^H)$	2
$\gamma_k=\gamma_{k-1}+1$	
$a_{n,k+1}=\pi_{n,k,k-1\|k}(\pi_{n,k-1\|k})^{-1}$	1
cost/update $21N$	

Talbe.2: Adaptive Component-Wise EM-Kalman Algorithm