LOW COMPLEXITY BAYESIAN ADAPTIVE FILTERING WITH INDEPENDENT $AR(1)$ FILTER COEFFICIENT MODELS

*Tayeb Sadiki and Dirk T.M. Slock**

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Email: {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.

1. INTRODUCTION

In Bayesian Adaptive Filtering (BAF) [1], the evolution of filter coefficients is modeled as a stationary process. A simple choice for search process is 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)
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.

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

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

where X_k is a $N \times 1$ input compelx vector and N is the length of the filter. The error or noise term v_k is assumed to be zero-mean uncorrelated normally distributed noise vector with common covariance matrix R. The BF series H_k^0 is assumed to be of primary interest [1]; it is modelled as a first order multivariate process of the form

$$H_k^0 = AH_{k-1}^0 + w_k \tag{3}$$

where $E[W_k W_k^H] = Q$ is a 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

*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.

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 kind of random effects model for time varying, 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. One may regard the X_k^H as fixed design input vector which define the way we observe the components of the BF H_k^0. In this paper, we provide a convenient method for dealing with the incomplete data problem introduced by missing observations.

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 (3)-(4) [6]. 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 $\theta = (A, Q)$ 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 avaible then ML estimation of A reduces to applying to (3). The covariance elements, Q, 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 $Z = (H_k^0, Y_k)$.

In order to develop a procedure for estimating the parameters in the state-space model defined by (5) and (6), we note first that the joint log-likelihood of the complet data Z can be written in the form

First, by repeated application of Bayes Rule

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

where $f_Z(z,\theta)$ is the probability vdensity associated with Z and $f_{Z|Y=y}(z,\theta).f_Y(y;\theta)$ is the conditional probability density of Z given $Y = y$. Taking the logarithm on both sides of (4),

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

Define, for convenience

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

With this definition we can writ

$$\begin{aligned}\log L \;=\; & -2\log f_\theta(H_k^0, Y_M, \theta|Y_M) \\ & + M\log detQ + \\ & + \sum_{k=1}^{M} tr(H_k^0 - AH_{k-1}^0)Q^{-1}(H_k^0 - AH_{k-1}^0)^H \\ & + \sum_{k=1}^{M} tr(y_k - X_k^H H_{k-1}^0)R^{-1}(y_k - X_k^H H_{k-1}^0)^H \end{aligned} \qquad (6)$$

where $\log L$ is to be maximized with respect to parameters A and Q . Since 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 Y. That is, the estimated parameters at the $(k+1)-th$ iterate as the values A and Q which maximize

$$U(\theta, \hat{\theta}_k) = E_{\hat{\theta}_k}\{\log f_\theta(H_k^0, Y_M, \theta|Y_M)\} \qquad (7)$$

where $E_{\hat{\theta}_k}$ denotes the conditional expectaion relative to a density containing the $k\,th$ iterate values.

In order to calculate the conditional expectation defined in (7), it is convenient to define the conditional mean

$$\begin{aligned}\hat{H}_k^0 &= E_{\hat{\theta}_k}\{H_k^0|Y_M\} \\ \hat{P}_k &= E[\tilde{H}_k^0\tilde{H}_k^{0H}|Y_M\} \\ \hat{P}_{k-1} &= E[\tilde{H}_{k-1}^0\tilde{H}_{k-1}^{0H}|Y_M\}\end{aligned} \qquad (8)$$

we suppose the following definitions

$$\begin{aligned} B1 &= \sum_{k=1}^{M}(E_{\hat{\theta}_k}\{H_{k-1}^0(H_{k-1}^0)^H|Y_M\} + \hat{P}_{k-1}) \\ B2 &= \sum_{k=1}^{M}E_{\hat{\theta}_k}\{H_k^0(H_k^0)^H|Y_M\} + \hat{P}_k \\ B12 &= \sum_{k=1}^{M}E_{\hat{\theta}_k}\{H_k^0(H_{k-1}^0)^H|Y_M\} + \hat{P}_{k,k-1} \end{aligned} \qquad (9)$$

The Kalman filter terms $\hat{H}_k^0$, $\hat{P}_k$ and $\hat{P}_{k,k-1}$ are computed under the parameter values $A^{(k)}$ and $Q^{(k)}$ using the recursions in (7).Furthermore, it is easy to see that the choices

$$\mathbf{\hat{Q}} = \frac{1}{M}(B2_k - B12_k B1_k^{-1} B12_k^H) \qquad (10)$$

$$\mathbf{\hat{A}} = B12_k B1_k^{-1} \qquad (11)$$

maximize the last two lines in the likelihood function (6).

3. ADAPTIVE 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 for iteratively update the parameter model.
The algorithm is resulting in **Fig. 2**.

4. SIMULATION

The behavior of Adaptive Kalman and Kalman filters are compared on the basis of simulation results, as shown in Fig. 1. The proposed algorithms are implemented with the parameters $\alpha = 0.9$, $\lambda = 0.97$

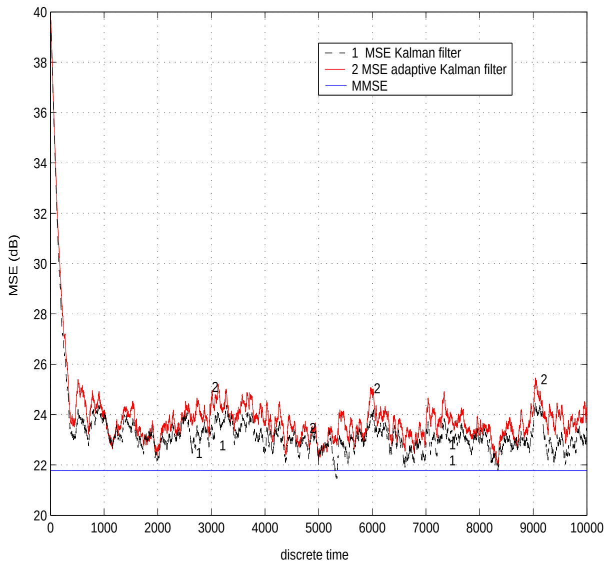

Fig. 1. Comparison between the proposed Adaptive Kalman algorithm and Kalman filter

5. 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.

6. 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 Sgnal Prossing Series, 1995.

[5] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," in *Journal of Time Series Analysis vol. 3, No. 4*, 1982.

[6] S. Haykin, *Adaptive Filter Theory (3rd Edition)*, Prentice Hall, 1996.

Adaptive Kalman	
Computation	**Cost** $(\times)$
Initialization	
$\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{0}\|\mathbf{0}}=\hat{\mathbf{0}}$, $\mathbf{P}_{\mathbf{0}\|\mathbf{0}}=\mathbf{100I}$,	
$\mathbf{A}^{(\mathbf{0})}=(\mathbf{1}-\alpha)\mathbf{I}$, $\mathbf{Q}^{(\mathbf{0})}=\alpha\mathbf{I}$	
$\mathbf{B_1}^{(\mathbf{0})}=\mathbf{0}$, $\mathbf{B_2}^{(\mathbf{0})}=\mathbf{0}$	$2N$
$\gamma^{(0)}=0$	
Kalman filtering and one step smoothing	
$\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l-1}}=\hat{\mathbf{A}}^{(\mathbf{l})}\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l-1}\|\mathbf{l-1}}$	N^2
$\hat{y}_{l\|l-1}=\mathbf{x_l^H}\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l-1}}$	N^2
$\mathbf{l_l}=\mathbf{P_{l\|l-1}x_l}(\mathbf{x_l^H P_{l\|l-1}x_l}+\sigma_{\mathbf{v}}^{\mathbf{2}})^{\mathbf{-1}}$	$3N^2+N+1$
$\mathbf{C_{l-1}}=\mathbf{P_{l-1\|l-1}}(\mathbf{A}^{(\mathbf{l})})^{\mathbf{H}}\mathbf{P}^{\mathbf{-1}}_{\mathbf{l-1\|l-1}}$	$2N^2$
$\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l}}=\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l-1}}\mathbf{l_l}(\mathbf{y_l}-\hat{\mathbf{y}}_{\mathbf{l}\|\mathbf{l-1}})$	$2N$
$\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l-1}\|\mathbf{l}}=\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l-1}\|\mathbf{l-1}}$	
$+\mathbf{C_{l-1}}(\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l}}-\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l-1}})$	N^2
$\mathbf{P_{l\|l}}=(\mathbf{I}-\mathbf{l_l x_l^H})\mathbf{P_{l\|l-1}}$	$2N^2$
$\mathbf{P_{l-1\|l}}=\mathbf{P_{l-1\|l-1}}+\mathbf{C_{l-1}}(\mathbf{P_{l\|l}}-\mathbf{P_{l\|l-1}})$	N^2
Model Paramters Adaptation	
$B_1^{(l)}=\lambda B_1^{(l-1)}+\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l}}\hat{\mathbf{H}}^{\mathbf{0}\ \mathbf{H}}_{\mathbf{l}\|\mathbf{l}}+\mathbf{P_{l\|l}}$	N^2
$B_2^{(l)}=\lambda B_2^{(l-1)}+\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l-1}\|\mathbf{l-1}}\hat{\mathbf{H}}^{\mathbf{0}\ \mathbf{H}}_{\mathbf{l-1}\|\mathbf{l-1}}+\mathbf{P_{l-1\|l}}$	N^2
$B_{12}^{(l)}=\lambda B_{12}^{(l-1)}+\hat{\mathbf{H}}^{\mathbf{0}}_{\mathbf{l}\|\mathbf{l}}\hat{\mathbf{H}}^{\mathbf{0}\ \mathbf{H}}_{\mathbf{l-1}\|\mathbf{l}}$	
$+\mathbf{P_{l\|l}C^H_{l-1}}$	$2N^2$
$\hat{\mathbf{Q}}^{(\mathbf{l+1})}=\frac{1}{\gamma^{(\mathbf{l})}}(\mathbf{B_1^{(l)}}-\mathbf{B_{12}^{(l)}}(\mathbf{B_2^{(l)}})^{\mathbf{-1}}+(\mathbf{B_{12}^{(l)}})^{\mathbf{H}})$	$4N$
$\gamma^{(l)}=\lambda\gamma^{(l-1)}+1$	
$\mathbf{A}^{(\mathbf{l+1})}=\mathbf{B_{12}^{(l)}}(\mathbf{B_2^{(l)}})^{\mathbf{-1}}$	$2N$
$\mathbf{P_{l+1\|l}}=\mathbf{A}^{(\mathbf{l+1})}\mathbf{P_{l\|l}}(\mathbf{A}^{(\mathbf{l+1})})^{\dagger}+\hat{\mathbf{Q}}^{(\mathbf{l+1})}$	$2N^2$
cost/update $16N^2+11N+1$	

Fig. 2.Adaptive Kalman Algorithm