A TWO-STAGE APPROACH TO BAYESIAN ADAPTIVE FILTERING

Tayeb Sadiki, Dirk T.M. Slock

Eurecom Institute
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Tel: +33 4 9300 8156/8106; Fax: +33 4 9300 8200
e-mail:{sadiki,slock}@eurecom.fr

ABSTRACT

The purpose of this paper is to introduce Bayesian Adaptive Filtering (BAF) techniques that are not immensely more complex than the LMS algorithm. The proposed two-stage approach consists of a first stage employing a basic fast tracking adaptive filter, followed by lowpass filtering and downsampling of the time-varying filter coefficients. The second stage then applies Kalman filtering at the reduced rate on a simplified state-space model, with an additive white noise measurement equation. The parameters in the state equation can be conveniently identified with an adaptive EM algorithm. The first stage would typically employ a (Normalized) LMS algorithm with a large stepsize. The main assumption underlying the proposed two-stage approach is that even in fast tracking applications, the bandwidth of the optimal filter variation is typically small compared to the signal bandwitdh, motivating the downsampling operation. The first stage attempts to provide a bias-free filter estimate whereas the second stage optimizes the estimation variance. The performance of the proposed scheme is evaluated by simulations.

1. INTRODUCTION

Adaptive filtering is essentially intended for tracking time-varying optimal filters. The time variation of the optimal filter can be described by either expanding the filter coefficients into fixed time-varying (e.g. sinusoidal) basis functions (basis expansion models (BEMs)) [1] or by modeling them as stationary processes. The latter approach is perhaps better suited for minimum delay online processing. This case of constant slow variation of the filter coefficients (”drifting” parameters) is to be contrasted with another possible case of only occasional but significant variation (”jumping” parameters) which shall not be considered here. A lot of work has been done on optimizing the single parameter regulating the tracking speed of classical LMS or exponentially weighted RLS algorithms [2],[3]. For LMS, such an adaptive optimization leads to the class of Variable Step-Size (VSS) algorithms, see e.g. [4] and references therein. Adaptive filtering algorithms with a single adaptation parameter do not take into account that different portions of the filter may have different variation speeds and/or different magnitudes and hence are quite suboptimal. One noteworthy attempt to overcome this limitation is the introduction of a coefficient-wise VSS, as in [5], but the automatic adaptation of these VSSs is a difficult task.

In Bayesian Adaptive Filtering (BAF), prior information on the filter coefficient variances and variation spectra is exploited to optimize adaptive filter performance. A straightforward way to implement BAF is to use the Kalman filter, see e.g. [6],[7]. However, the complexity of the Kalman filter is enormous compared to that of the popular LMS adaptive filtering algorithm. Furthermore, the Kalman filter needs to be augmented with a state-space model identification technique.

Consider now the prototype adaptive filtering set-up, which is the system identification set-up, in which the desired-response signal d_k is modeled as the output of the optimal filter, which can be time-varying, plus independent (white) noise:

$$d_k = X_k^H H_k + v_k \tag{1}$$

where $X_k^H = [x_k \; x_{k-1} \cdots x_{k-N+1}]$ is the input signal vector and all terms are complex-valued. The input vector X_k is known up to time k and is assumed stationary with zero mean and nonsingular covariance matrix $R = E[X_k X_k^H]$ Our aim is to estimate the time-varying parameter column vector H_k. Some general references on the tracking behavior of adaptive filtering algorithms are [2], [3],[8],. In this work we consider Bayesian Adaptive filtering based on a two-stage approach. A first stage with a fast standard adaptive filter, e.g. NLMS with stepsize equal to one. After some possible downsampling then, we consider an optimal filter

Eurécom’s research is partially supported by its industrial partners: BMW, Bouygues Telecom, Cisco Systems, France Télécom, Hitachi Europe, SFR, Sharp, ST Microelectronics, Swisscom, Thales. The research reported herin was also partially sponsored by the European Network of Excellence NewCom and by a PACA regional scholarship.

in the second stage to extract H_k from the NLMS estimates, see figure (1).

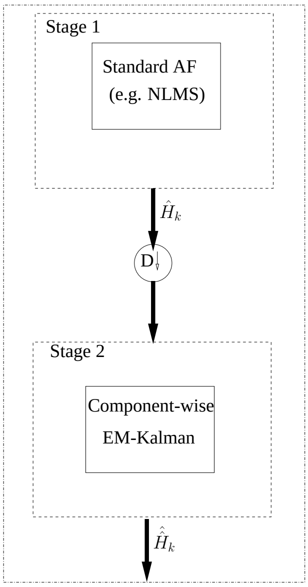

Fig. 1. Two-stage adaptive filtering.

2. STAGE 1: NLMS ALGORITHM

The simplest choice for a fast converging adaptive filtering (AFing) algorithm is a LMS algorithm with large stepsize, preferably the NLMS algorithm with normalized stepsize equal to one, or a smaller value of that order of magnitude. The NLMS algorithm updates the adaptive filter coefficients according to

$$e_k = d_k - X_k^H \hat{H}_{k-1} \quad (2)$$

$$\hat{H}_k = \hat{H}_{k-1} + \frac{\mu}{\|X_k\|^2} X_k e_k^* \quad (3)$$

For colored input and a FIR filter of length N, NLMS converges in general with N different modes that are of the form [9]

$$\frac{1}{1 - \sqrt{1 - \mu(2-\mu)\,\frac{\lambda_i}{\operatorname{tr} R}}\, q^{-1}} \quad (4)$$

where μ is the NLMS stepsize, we have assumed $N \gg 1$, the λ_i are the eigenvalues of the input signal covariance matrix $R = R_{XX}$, and q^{-1} is the one sample delay operator: $q^{-1} x_k = x_{k-1}$. We shall call the variation bandwidth of the optimal filter the Doppler bandwidth, which is the customary terminology when the adaptive filter represents a wireless channel response. We are going to assume that the Doppler bandwidth is much smaller than the signal bandwidth. For simplicity, let us assume that the input signal is not too colored so that the Doppler bandwidth can be smaller than the bandwidth $f_i = \frac{\mu(2-\mu)}{2\pi}\,\frac{\lambda_i}{\operatorname{tr} R}$ of each of the eigenmodes. In this case, the NLMS adaptive filtering algorithm will pass the optimal filter coefficients undistortedly (zero bias). It will only introduce an estimation noise. In steady-state, this estimation noise leads to an estimation error $\widetilde{H}_k = H_k - \widehat{H}_k$ with covariance matrix $R_{\widetilde{H}\widetilde{H}} = \dfrac{\mu}{2-\mu}\,\dfrac{\sigma_v^2}{\operatorname{tr} R}\, I_N$. So the errors on the various filter components are uncorrelated and of identical variance. The errors are not temporally white however, due to the coloring introduced by the filtering of the modes in (4). However, due to the previous assumptions, the estimation noise can be considered white over the Doppler bandwidth of the optimal filter.

A better alternative to the NLMS algorithm in the first stage would be an adaptive filter that is less sensitive to the input signal color. If we want no such sensitivity then a Recursive Least-Squares (RLS) algorithm should be used. To minimize the distortion (so-called ”lag noise”) on the optimal filter the best RLS choice would be one with a sliding rectangular window, in which a delay gets introduced equal to half the window length (non-causal adaptive filtering) [10]. RLS algorithms are more complex that (N)LMS, but fast versions exist. There is also a whole range of adaptive filtering algorithms in between LMS and RLS in terms of complexity and performance, such Affine Projection Algorithms, Fast Newton Transversal Filters, frequency domain adaptive filters, LMS with prewhitening etc.

3. SUBSAMPLING GLUE

As mentioned earlier, if the Doppler bandwidth is significantly less than the signal bandwidth (sampling rate), then it would be overkill to put in place an optimal tracking algorithm working at the sampling rate. In that case, the output of the first stage (the vector sequence $\widehat{H}_k$) can be lowpass filtered and commensurate downsampled without introducing distortion (lag noise) as long as the lowpass filter does not distort the Doppler spectrum. The main goal of this operation is to reduce complexity. Indeed further processing in the second stage can now be performed at a reduced rate. And fixed lowpass filtering does not have to be a complex operation (if a simple filter is used, for instance first order IIR (exponential averaging)). Another reason is that, whereas it would constitute quite an approximation to model $\widetilde{H}_k$ as temporally white, after lowpass filtering and downsampling (with a factor D), such an approximation becomes more accurate. The lowpass filtering operation reduces the estimation noise roughly with a factro D. In what follows, we shall continue to use the same notation for the subsampled rate and continue to denote the lowpass filtered and subsampled NLMS output as $\widehat{H}_k$. This provides the measurement data for stage two.

4. STAGE 2: "DIAGONAL" EM-KALMAN FILTERING

Consider the state-space model

$$\begin{aligned} H_{k+1} &= AH_k + W_k &(5)\\ \hat{H}_k &= H_k + \tilde{H}_k &(6)\end{aligned}$$

The measurement and process noise terms are assumed to be zero mean Gaussian with covariances $R_{\tilde{H}}$ and Q respectively.The matrix A contains information about how the states evolve. It is particularly useful in tracking applications. The matrix A should be viewed as a mechanism to achieve directed trajectories in state space. In other words, A allows for more general jumps than the simple random walk that would result by excluding A from the model. Despite the fact that the data is processed in batches, the model of equation (6) allows the weights to be time varying. It is, therefore, possible to deal with non-stationary data sets. In the event of the data being stationary, we should expect the process noise term to vanish. Consequently, if we know that the data is stationary, the estimate of the process noise can be used to determine how well the model explains the data. The objective is to estimate the model states (weights)H_k and the set of parameters $\phi = \{A, Q, R\}$ given the measurements $\hat{H}_{1:N}$. Then we use a Kalman smoother to estimate H_k and EM algorithm to estimate the set of parameters. Since the Kalman model is diagonal, we propose a Component-Wise Adaptive Kalman algorithm to update the filter coefficients, which decreases computational complexity. Then the model (6) becomes:

$$\begin{aligned} h_{i+1} &= a\,h_i + w_i &(7)\\ \hat{h}_i &= h_i + \tilde{h}_i &(8)\end{aligned}$$

4.1. Kalman smoother

Smoothing often entails forward and backward filtering over a segment of data so as to obtain improved averaged estimates. Various techniques have been proposed to accomplish this goal . This study uses the well-known Rauch-Tung-Striebel smoother . The forward filtering stage involves computing the estimates $\hat{h}_k$ and P_k , over a segment of I samples, with the following KF recursions:

$$\begin{aligned} \hat{\hat{h}}_{i+1|i} &= a\hat{\hat{h}}_i &(9)\\ p_{i+1|i} &= aa^* p_i + q &(10)\\ k^f_{i+1} &= p_{i+1|i}(r + p_{i+1|i})^{-1} &(11)\\ \hat{\hat{h}}_{i+1} &= \hat{\hat{h}}_{i+1|i} + k^f_{i+1}(\hat{h}_{i+1} - \hat{\hat{h}}_{i+1|i}) &(12)\end{aligned}$$

where k^f denotes the Kalman gain . Subsequently, the Rauch-Tung-Striebel smoother makes use of the following backward recursions:

$$\begin{aligned} J_{i-1} &= \frac{p_{i-1}a^*}{p_{i|i-1}} &(13)\\ \hat{\hat{h}}_{i-1|n} &= \hat{\hat{h}}_{i-1} J_{i-1}(\hat{\hat{h}}_{i|n} - a\hat{\hat{h}}_{i-1}) &(14)\\ p_{i-1|n} &= p_{i-1} + J_{i-1}(p_{i|n} - p_{i|i-1})J^*_{i-1} &(15)\\ p_{i,i-1|n} &= p_i J^*_{i-1} + J_i(p_{i+i|n} - ap_i)J^*_{i-1} &(16)\end{aligned}$$

where the parameters, covariance and cross-covariance are defined as follows:

$$\begin{aligned} \hat{\hat{h}}_{i+1|n} &= E[h_{i+1} \mid \hat{h}_{1:n}]\\ p_{i|n} &= E[(h_i - \hat{\hat{h}}_i)(h_i - \hat{\hat{h}}_i)^* \mid \hat{h}_{1:n}]\\ p_{i,i-1|n} &= E[(h_i - \hat{\hat{h}}_i)(h_{i-1} - \hat{\hat{h}}_{i-1})^* \mid \hat{h}_{1:n}] \end{aligned} \quad (17)$$

4.2. EM Algorithm

Due to lack of space we shall limit the discussion to an explanation with words. The EM algorithm iterates between an E step and an M step. In the E(stimation) step, the state estimates and state estimation error covariances are determined recursively using tha Kalman filter equations, using the state model values of the previous iteration. The M(aximization) (of the likelihood) step then essentially performs first-order linear prediction on the estimated state covariances to determine the prediction coefficients A and the prediction error covariance Q. Due to orthogonality property of linear MMSE estimation the state covariances (at lag 0 and 1) are the sum of the (sample) covariances of the state estimates plus the state estimation error covariances, both quantities being produced by the Kalman filter. To make the EM algorithm adaptive, we shall perform one iteration per time update. The fixed-interval smoothing gets transformed into fixed-lag smoothing. As the state model is AR(1), a lag of one sample turns out to be sufficient (a positive lag is nevertheless required, otherwise a chicken-and-egg problem arises; hence pure Kalman filtering, without some smoothing, does not work). See [11] for a derivation.

4.3. Model parameters adaptation

The state model parameters can be adapted according to

$$\begin{aligned} \psi_{i|i} &= \lambda\psi_{i,n|i-1} + (\hat{\hat{h}}_{i|i}\hat{\hat{h}}^*_{i|i} + p_{i|i})\\ \psi_{i-1|i} &= \lambda\psi_{i-1|i-1} + (\hat{\hat{h}}_{i-1|i}\hat{\hat{h}}^*_{i-1|i} + p_{i-1|i})\\ d_i &= p_{i|i}C^*_{i-1}\\ &= a_i p_{i-1|i-1} - k^f_i(a_i^{-1}(1 - q_i x_i))^* \end{aligned} \quad (18)$$

$$
\begin{aligned}
\psi_{i,i-1|i} &= \lambda\psi_{i,i-1|i-1} + (\hat{\hat{h}}_{i|i}\hat{\hat{h}}^*_{i-1|i} + d_i) \\
q_{i+1} &= \frac{1}{\gamma_i}(\psi_{i,|i} - \frac{\psi_{i,i-1|i}}{(\psi_{i-1|i})}(\psi_{i,i-1|i})^* \\
a_{i+1} &= \psi_{i,i-1|i}(\psi_{i-1|i})^{-1}
\end{aligned} \tag{19}
$$

5. NUMERICAL RESULTS

The behavior of two-stage adaptive algorithm and the NLMS algorithm are compared on the basis of simulation results, as shown in Fig. 2. The concept of the two-stage adaptive algorithm that we introduced earlier, based on modeling the optimal adaptive filter coefficients as a stationary vector process. The optimal parameters are $A = 0.95 * I$,where I is identity matrix, and the error covaraince matrix Q is an exponential power delay profile, with the characteristic parameter $\beta = 0.9$.

6. CONCLUSION

As Fig. 2 shows, the proposed two-stage adaptive algorithm converges to the MMSE. The convergence speed of the proposed is beter than NLMS algorithm and is comparable to that of the ideal Kalman filter (known parameters).

7. REFERENCES

[1] M. Niedzwiecki and T. Klaput. "Fast Recursive Basis Function Estimators for Identification of Time-Varying Processes". *IEEE Trans. on Signal Process.*, 50(8), Aug. 2002.

[2] S. Haykin. *"Adaptive Filter Theory"*. Prentice Hall, 2001. 4th edition.

[3] M. Niedzwiecki. *"Identification of Time-Varying Systems"*. Wiley, 2000.

[4] C. Rusu and C.F.N. Cowan. "Convex Variable Step-Size (CVSS) Algorithm". *IEEE Signal Proc. Letters*, 7(9), Sept. 2000.

[5] W. Liu. "Performance of Joint Data and Channel Estimation Using Tap Variable Step-Size (TVSS) LMS for Multipath Fast Fading Channel". In *Proc. Globecom*, pages 973–978, 1994.

[6] T. Sadiki and D.T.M. Slock. "Bayesian Adaptive Filtering: Principles and Practical Approaches". In *Proc. 12th European Sig. Proc. Conf. (EUSIPCO)*, Vienna, Austria, Sept. 2004.

[7] S. Haykin, A.H. Sayed, J.R. Zeidler, P. Yee, and P.C. Wei. "Adaptive Tracking of Linear Time-Variant Systems by Extended RLS Algorithms". *IEEE Trans. on Signal Process.*, 45(5):1118–1128, May 1997.

[8] L. Ljung L. Guo. "Performance Analysis of the Forgetting Factor RLS Algorithm". *Int. J. Adaptive Cont. Sig. Proces.*, 7:141–537, 1993.

[9] D.T.M. Slock. "On the Convergence Behavior of the LMS and the Normalized LMS Algorithms". *IEEE Trans. on Signal Processing*, 41(9):2811–2825, Sept. 1993.

[10] T. Sadiki, M. Triki, and D.T.M. Slock. "Window Optimization Issues in Recursive Least-Squares Adaptive Filtering and Tracking". In *Proc. 38th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2004.

[11] T. Sadiki and D.T.M. Slock. "Bayesian Adaptive Filtering At linear Cost". In *Proc. IEEE-SP Workshop Statistical Sig. Proc. (SSP)*, Bordeaux, France, July 2005.

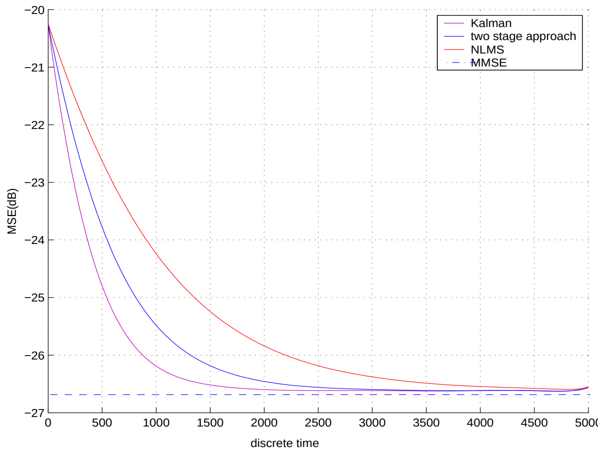

Fig. 2. Comparison between the proposed two-stage adaptive filter, NLMS and the Kalman filter with known optimal parameters.

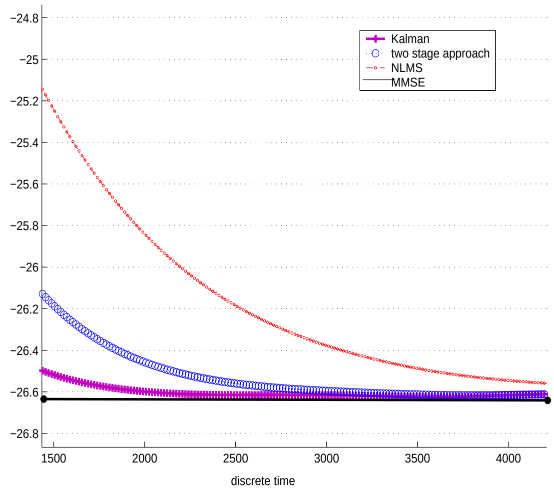

Fig. 3. Zoom on the steady-state behavior.