STEADY-STATE PERFORMANCE COMPARISON OF BAYESIAN AND STANDARD 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

It has been known for a long time that for best tracking results adaptive filtering should be formulated as a Kalman filtering problem, leading to Bayesian Adaptive Filtering (BAF). BAF techniques with acceptable complexity can be obtained by focusing on a diagonal AR(1) model for the time-varying optimal filter settings. The hyper-parameters of the AR(1) model can be adapted by introducing EM techniques and one sample fixed-lag smoothing at little extra cost. Standard AF techniques such as the LMS and RLS algorithms are equipped with only one hyper-parameter (stepsize, forgetting factor) to optimize their tracking behavior. In this paper we compare the steady-state tracking performance of Bayesian and standard AF techniques.

Index Terms*——— *Bayesian Adaptive Filter (BAF), LMS, RLS and Kalman algorithms, Tracking Ability, Time-varying system, Steady state analysis

1. INTRODUCTION

Adaptive filtering have been extensively studied for a large range of applications including channel estimation, adaptive equalization, echocancelation, etc in a variety of stationary environment. For the nonstationary environments, two different classes of input have been studied for adaptive filtering algorithms. It has been shown that the Wiener solution has a time-varying characteristic. In contrast to adaptive filter convergence, which is a transient phenomenon, the tracking characteristics of the adaptive filter are sknown be to a steady state property of the filter. Consequently, good convergence properties do not ensure good tracking performance, and a compromise between the two properties are required for applications in a non-stationary environment. The standard adaptive filtering (SAF) such as the least mean-square (LMS) algorithm , and the recursive least-squares (RLS) algorithm are established as the principal algorithms to track for linear adaptive filtering. The convergence behaviors of both of these algorithms can be found in the literature ([1]), ([2]), ([3]). The RLS algorithm has a faster rate of convergence than the LMS algorithm and is not sensitive to variations in the eigenvalues of correlation matrix of the input signal. However, when operating in a non-stationary environment, they pocess only one parameter to adjust the tracking. Most of the work on adapting trackive capability has focused on adapting one tracking parameter. In RLS, it doe cost any computational complexity to make the forgetting factor time-varying. Modifications to fast RLS algorithms to allow a time-varying forgetting factor, as well as algorithms to adjust this forgetting factor on the basis of correlation matching have been pursued in [4]. The equivalent development for LMS algorithms concerns Variable Step-Size (VSS) algorithms. Important developments were presented in [5],[6],[7],[8],[9] [10],[11]. Most of the VSS algorithms use the steepest-descent strategy and the instantaneous squared error cost function of the LMS algorithm to adjust the additional parameter, which is the step-size. A related but different approach consists in running various adaptive filters with different time constants and selecting or combining their outputs, similarly to what is done in model order selection, see [12],[13],[14],[15].

A further refinement is to allow different tracking bandwidths for different filter components as is done in [16] with a VSS per filter coefficient and in [17] where the tracking capacity increases with frequency for the various frequency domain components of the filter. The work in [16] essentially shows that a "diagonal" state-space model may allow a simplification of the Kalman Filter (KF) to a LMS algorithm with a VSS per tap, but no attempt is made to automatically adjust the resulting stepsizes.

Besides the statistical modeling of the parameter variation, another important ingredient in Bayesian adaptive filtering is the incorporation of prior knowledge on the coefficient sizes. The influence of the prior distribution on Bayesian estimation, depends on the confidence on the observation, which in turn depends on the length of the observation, and on the SNR. In general, as the number of the observation samples and the SNR increase, the variance of the estimate, and the influence of the prior, decrease. In estimating a Gaussian distributed parameter observed in AWGN, as the length of the observation N increases, the importance of the prior decrease, and the MAP estimate tends to the ML estimate.

Indeed, when tracking time-varying filters, it becomes possible to learn the variances of the filter coefficients. This aspect has been exploited for a while in a rudimentary, binary form for sparse filters: filter coefficients are either adapted or deemed to small and kept zero (for each filter coefficient, the stepsize is either 0 or a constant). More recently, a smoother evolution of the stepsize has been introduced, leading to the Proportionate LMS (PLMS) algorithm, motivated e.g. by acoustic echo cancellation in which the adaptive filter has many coefficients, but their value tapers off, see [18],[19].

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.

Similar prior information is starting to be taken into account for (LMMSE) channel estimation in wireless communications [20], where the evolution of the channel coefficient variances along the impulse response is called the power delay profile.

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 [3], [21] 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 [22],[23],. For LMS, such an adaptive optimization leads to the class of Variable Step-Size (VSS) algorithms, see e.g. [24] 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 can be quite suboptimal. One noteworthy attempt to overcome this limitation is the introduction of a coefficient-wise VSS, but the automatic adaptation of these VSSs is a difficult task. In Bayesian Adaptive Filtering (BAF)[25], 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. However, the complexity of the Kalman filter is much higher 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.
A non-stationary process can be defined as one whose statistical parameters are a time-varying. In our case the state control process model can be described as an AR(1) The hyperparameters of the AR(1) model can be adapted by introducing EM techniques and one sample fixed-lag smoothing at little extra cost [25] what carries out has an optimal solution (Kalman filter), leading to Bayesian approach, with special consideration for the complexities. The comparison between BAF and standard AF is done in terms of the steady-state excess mean-square estimation error (EMSE). The remainder of this paper is organized as fellows.
Section II, we develop the Kalman filter on the Bayesian point of view, and establishes the relevant notation that will be used in the sequel. Section III, we derives a new analytical expression of the excess mean square error (EMSE) of BAF and SAF . Section IV presents experimental results of computer simulations that verify the developed theory. Finally, section V is a brief conclusion.

2. KALMAN FILTER IN A BAYESIAN POINT OF VIEW

Consider now the prototype adaptive filtering set-up, which is the system identification set-up, in which the desired response signal y_k is modeled as the output of the optimal filter, which can be time-varying, plus independent (white) noise. The adaptive system identification Fig. **??** is designed for determining a (typically linear FIR) model of the transfer function for an unknown, time-varying digital or analog system. The time-varying optimal filter coefficients is given by the equation:

$$H_k = A\,H_{k-1} + W_k \qquad (1)$$

where W_k is a random vector of size N of covariance Q. The complexity of Kalman filter can be limited to $O(N^2)$ by taking A and Q diagonal, the adaptive filter order, in physical terms, the tap-weight vector H_k may be viewed as originating from the noise process W_k, whose individual elements are applied to a bank one-pole low-pass filters. Each such filter has a transfer function equal to $1/(1-A_i q^{-1})$ where q^{-1} is the unit-delay operator and A_i is the i^{th} diagonal element of A. It is assumed that $A \leqslant I$ where I is identity matrix , the significance of this assumption is that the bandwidth of the low pass filters is very much smaller than the incoming data rate.

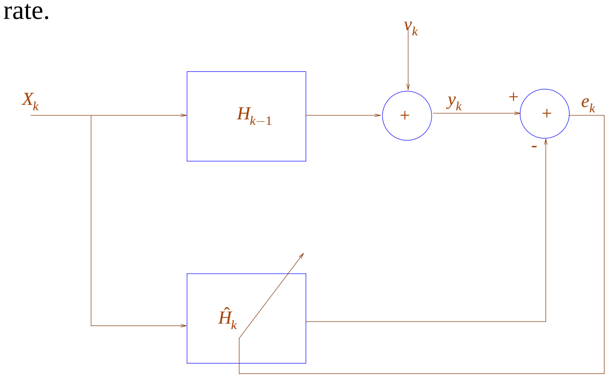

Figure 1: System identification block diagram

Let consider the system driven by noise, with noisy observation

$$\begin{aligned} H_k &= AH_{k-1} + W_k \\ y_k &= X_k^H H_{k-1} + v_k \end{aligned} \qquad (2)$$

where $X_k^H = [x_k x_{k-1} \ldots x_{k-N+1}]$ is the input signal vector . The input vector X_k is known up to time k and is assumed stationary with zero mean and nonsingular covariance matrix $R = E[X_k X_k^H]$. In this section we derive the Kalman filter from the Bayesian point of view. For the Bayesian approach, we assume that the noise processes are Gaussian distributed. Then the Bayes estimate of H_k amounts to finding the conditional mean of H_k, given the observations.
The key equation in the Bayes derivation is the time update step,

$$f(H_k \mid Y_k) = \int f(H_k \mid H_{k-1}) f(H_{k-1} \mid Y_k) dH_{k-1} \qquad (3)$$

from which the estimate is propagated using the state update equation into the future; and

$$\begin{aligned} \underbrace{f(H_k \mid Y_{k+1})}_{posterior} &= \frac{f(y_{k+1} \mid H_k)}{f(y_{k+1} \mid Y_k)} \underbrace{f(H_k \mid Y_k)}_{prior} \\ Y_k &= [y_k, \ldots, y_{k-M}] \end{aligned} \qquad (4)$$

which is the measurement update step.
We will begin by finding explicit formulas for the time-update in (3).

1. The density $f(H_{k-1} \mid Y_k)$ corresponds to the estimate of H_{k-1}, given the measurements up to time k. Under the assumption and using the notation just introduced, the random variable H_{k-1} conditional upon Y_k is Gaussian,

$$H_{k-1} \mid Y_k \sim N(\hat{H}_{k-1|k-1}, P_{k-1|k-1}) \qquad (5)$$

2. The density $f(H_k \mid H_{k-1})$ is obtained by noting from (2) that, conditional upon H_{k-1}, H_k is distributed as

$$H_k \mid H_{k-1} \sim N(AH_{k-1}, Q) \qquad (6)$$

Inserting (5) and (6) into (3) and performing the integration (which involves expanding and completing the square), we find that $H_k \mid Y_k$ is Gaussian, with mean

$$\hat{H}_{k|k-1} = A\hat{H}_{k-1|k-1} \qquad (7)$$

and the error covariance is given by

$$P_{k|k-1} = AP_{k-1|k-1}A^H + Q \qquad (8)$$

The estimate of the state is updated using the following steps (for mor details see the appendix)

$$\begin{aligned}
\hat{H}_{k|k-1} &= A\hat{H}_{k-1|k-1}, \\
P_{k|k-1} &= AP_{k-1|k-1}A^H + Q, \\
K_k^f &= P_{k|k-1}X_k(X_k^H P_{k|k-1} X_k + \sigma_v^2)^{-1}, \\
\hat{H}_{k|k} &= \hat{H}_{k|k-1} + K_k^f(y_k - X_k^H \hat{H}_{k|k-1}), \\
P_{k|k} &= (I - K_k^f X_k^H)P_{k|k-1}. \qquad (9)
\end{aligned}$$

3. STEADY-STATE EXCESS MEAN-SQUARE ERROR (EMSE)

Approximate analysis by RLS tracking analysis, assuming relatively slow variation see (2):

$$P \approx E[P] \; and \; E[P^{-1}] \approx (E[P])^{-1}$$

is still difficult if Q, R and A do not have the same eigenvectors. In our model A is diagonal, since Q is diagonal, in order for R to have the same eigenvectors need R be diagonal then $R = \sigma_x^2 * I$, with I is identity. Then P will be diagonal if initialized with diagonal matrix, in any case, P diagonal in steady-state. The time constant and the power delay profile are givens by $\tau = \frac{1}{1-|A_i|^2}$ and $Q_i = Q_1\beta^i$ respectively. The state estimate update is given by Kalman as :

$$\begin{aligned}
\hat{H}_{k|k} &= A\hat{H}_{k-1|k-1} + K_k(y_k - X_k A\hat{H}_{k-1|k-1}) \\
&= A\hat{H}_{k-1|k-1} + K_k X^H (H_{k-1} - A\hat{H}_{k-1|k-1}) \\
&\quad + K_k e_{opt} \\
&= A\hat{H}_{k-1|k-1} \\
&\quad + \frac{P_{k|k} X_k X_k^H}{\sigma_v^2}(H_{k-1} - A\hat{H}_{k-1|k-1}) \\
&\quad + K_k e_{opt} \qquad (10)
\end{aligned}$$

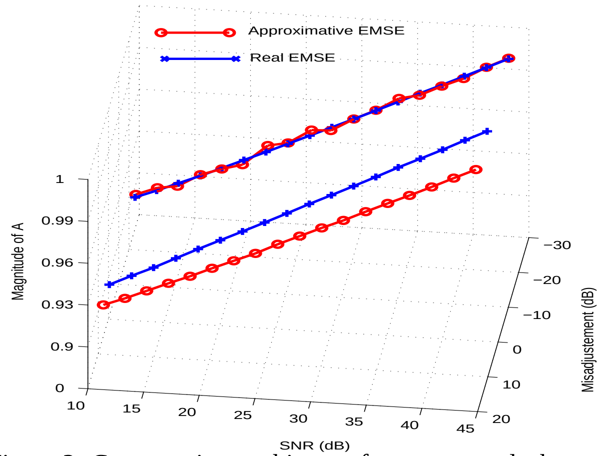

Figure 2: Comparative tracking performance results between misadjustement given by exact and approximate EMSEs for slow ($A_i = 0.99$ $\tau = 5N$) and Medium ($A_i = 0.90$ $\tau = 0.5N$) variations at SNR = 15 dB and $\beta = 0.9$

Where $K_k = \frac{P_{k|k} X_k}{\sigma_v^2}$ and e_{opt} represents the minimum(in a mean square sense) error at time k. In studying tracking behavior, we may exclude the influence of the estimation noise, since the deviation of $E[H_{k|k}]$ from H_k determines the response of the BAF algorithm to the non-stationarity of the environment. Taking expected values on both sides of (10), we get

$$\begin{aligned}
E[\hat{H}_{k|k}] &= AE[\hat{H}_{k-1|k-1}] \\
&\quad + \frac{P_{k|k} E[X_k X_k^H]}{\sigma_v^2})(H_{k-1} - AE[\hat{H}_{k-1|k-1}]) \\
&= AE[\hat{H}_{k-1|k-1}] \\
&\quad + \frac{P_{k|k} R}{\sigma_v^2}(H_{k-1} - AE[\hat{H}_{k-1|k-1}]) \qquad (11)
\end{aligned}$$

the lag-error is given by

$$\tilde{H}_k = E[\hat{H}_{k|k}] - H_k$$

$$\begin{aligned}
\tilde{H}_k &= AE[\hat{H}_{k-1|k-1}] - \frac{P_{k|k} RA}{\sigma_v^2}(\tilde{H}_{k-1}) \\
&\quad + \frac{P_{k|k} R}{\sigma_v^2}(H_{k-1} - AH_{k-1}) - H_k + AH_{k-1} \\
&= AE[\hat{H}_{k-1|k-1}] - \frac{P_{k|k} RA}{\sigma_v^2}(\tilde{H}_{k-1}) \\
&\quad + (\frac{P_{k|k} R}{\sigma_v^2}(I - A) + A)H_{k-1} - H_k \\
&\approx (I - \frac{P_{k|k} R}{\sigma_v^2})A\tilde{H}_{k-1} + AH_{k-1} - H_k \qquad (12)
\end{aligned}$$

The input is considered to be white ($R = \sigma_x^2 I$), note that, each element of the lag-error vector is determined by the following relation:

$$\tilde{h}_i(k) = (1 - \frac{p_{i,k|k}\sigma_x^2}{\sigma_v^2})A_i \tilde{h}_{i,k-1} + A_i h_{i,k-1} - h_{i,k}. \qquad (13)$$

where $\tilde{h}_i(k)$ is the i^{th} element of $\tilde{H}_k$. By properly interpreting the equation above, we can say that the lag is generated by applying the transformed instantaneous optimal coefficient to a first-order discrete-time filter denoted ***lag filter***

$$\tilde{H}_i(z) = \frac{A_i z^{-1} - 1}{1 - (1 - \frac{\sigma_x^2 p_{i,k|k}}{\sigma_v^2}) A_i z^{-1}} H_i(z) \qquad (14)$$

Using the inverse z-transform, the variance of the elements of the vector $\tilde{H}(k)$ can then be calculated by

$$E[\tilde{h}_i(k)\tilde{h}_i^H(k)] = \frac{1}{2\pi j}\oint \tilde{H}_i(z)\tilde{H}_i(z^{-1})Q_i z^{-1} dz$$

The BAF excess mean square error due to lag is then given

by Eq. (18)

$$EMSE = \sum_{i=1}^{N} EMSE_i \qquad (15)$$

where

$$EMSE_i = \sigma_x^2 p_i = \sigma_x^2 \frac{-(\frac{Q_i}{SNR_i} + Q_i) + Q_i\sqrt{(1+\frac{1}{SNR_i})^2 + 4\frac{\sigma_x^2}{\sigma_v^2} \mid A_i \mid^2}}{2 \mid A_i \mid^2} \qquad (16)$$

where $SNR_i = \frac{\sigma_x^2 Q_i}{\sigma_v^2(1-|A_i|^2)}$. The SA excess mean square error due to lag is then given by Eq. (18)

$$\begin{aligned} EMSE &= E[\tilde{H}(k)R\tilde{H}^H(k)] \qquad (17)\\ &= E[tr(\tilde{H}(k)R\tilde{H}^H(k))] \\ EMSE^{RLS} &= N\sigma_v^2 \frac{(1-\lambda)}{1-\lambda} + \frac{\sigma_x^2}{2}\sum_{i=1}^{N} \frac{A_i(1-Q_i)}{A_i - \lambda Q_i} \end{aligned}$$

The EMSE for LMS is obtained by using $\mu = \frac{1-\lambda}{1+\lambda}$.

where λ and μ are the forgetting factor and stepsize respectively. For an optimum value of the forgetting factor and stepsize the RLS and LMS are the same, then $EMSE_{opt}^{lms} = EMSE_{opt}^{rls}$.

3.1 Simplified Expression for Bayesian Adaptive Filtering

In simplified scenarios (e.g. high SNR, slow variation) we can write

$$\begin{aligned} p_i &= \frac{1 - \mid A_i \mid^2}{1 + SNR_i} \\ &= \frac{4\pi\beta_i}{1 + SNR_i} \end{aligned}$$

where $1 - \mid A_i \mid^2 = 4\pi\beta_i$
The misadjustement of BAF is given by

$$\begin{aligned} M &= \frac{EMSE}{MMSE} \\ &= 2\pi\frac{\sigma_x^2}{\sigma_v^2}\sum_{i=1}^{N}\frac{\beta_i}{1+SNR_i} \end{aligned}$$

If all β_i are the same, then $\sum_{i=1}^{N}\frac{1}{1+SNR_i}$ is the inverse of the harmonic mean of $1 + SNR_i$ and $\frac{1}{(\sum_{i=1}^{N}\frac{1}{1+SNR_i})} \leq \sum_{i=1}^{N}(1 + SNR_i)$ (the arithmetic mean) And β_i can be lows if A_i is complex $A_i = \mid A_i \mid e^{j2\pi f_i}$, where f_i is Doppler shift and $\mid A_i \mid = 1 - 2\pi\beta_i$ (if $\mid A_i \mid$ is sufficiently close to one) Then in this case the EMSE of Kalman, depend only on the PDP β_i, while the EMSE in the SAF case depend on the PDP and f_i.

4. NUMERICAL RESULTS

In this section the behavior of BAF and standard adaptive filters are compared for non-stationary environment in a system identification setup. The performance of the both algorithms are selected to produce a comparable level of misadjustement. In all simulations presented here, the desired signal y_k is corrupted by zero mean, (*iid*) Gaussian noise of σ_v^2 variance.
The proposed algorithms are implemented with the model parameters $A_i = 1 - \alpha/N$, with $\alpha = 0.4$, Q_i is chosen such as $\frac{Q_N}{Q_1} \ll 1$, the length of BAF is $N = 20$. In Fig. 3, the total excess mean square error ($EMSE$) has been plotted for a BAF and SAF for SNR varying from 10 to 48 and the input signal is assumed to be white . ‘

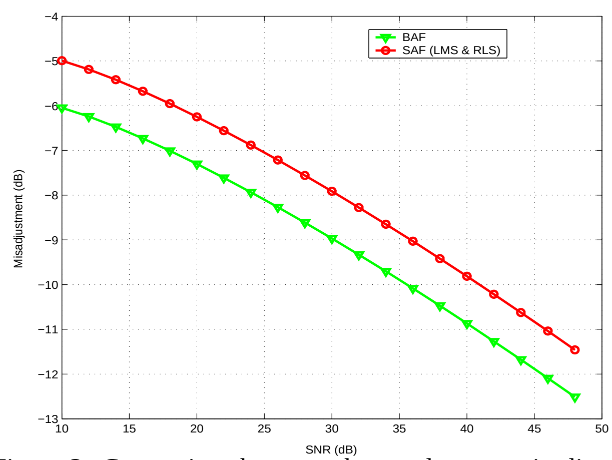

Figure 3: Comparison between the steady-state misadjustement of BAF and Standard AF

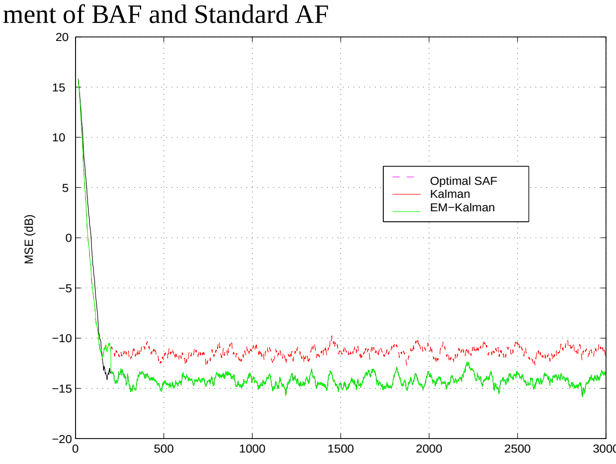

Figure 4: Comparative tracking performance results between BAF and Standard AF using MSE at SNR = 15dB

5. CONCLUSION

In this paper, we studied the steady-state performance of BAF and standard AF algorithms with optimized parameters. Analytical expressions for the steady-state excess mean-square error (EMSE) were calculated and verified by computer simulations. Results show that the BAF algorithm over those algorithms in a non-stationary environment in terms of misadjustement in both high and low SNR.

As Fig. 3 show, the BAF given clearly outperforms than the classical approaches for different SNR. Fig. 4 show, the BAF is close to Kalman filter with known parameters.

- Accurate approximate steady-state analysis of BAF provides simple expressions in simplified scenarios (e.g. high SNR, slow variation)
- BAF allows significant performance gains over RLS at comparable complexity.

REFERENCES

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

[2] B. Widrow *et al.* "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter". *Proc. IEEE*, 64(8):1151–1162, Aug. 1976.

[3] 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.

[4] D.T.M. Slock. "Fast Transversal Filters with Data Sequence Weighting". *IEEE Trans. on Acoustics, Speech and Signal Processing*, 37(3):346–359, March 1989.

[5] R.K. Raymond and W.J. Edward. "A Variable Step Size LMS Algorithm". In *IEEE Trans. on Signal Proc.*, volume 40, July 1992.

[6] J.B. Evans, P. Xue, and B. Liu. "Analysis and Implementation of Variable Step Size Adaptive Algorithms". *IEEE Trans. Signal Proc.*, 41(8):2517–2535, Aug. 1993.

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

[8] S.B. Gelfand, Y. Wei, and V.K. James. "The Stability of Variable Step-Size LMS Algorithms". *IEEE Trans. on Signal Proc.*, 47.

[9] R.C. Bilcu, P. Kuosmanen, and K. Egiazarian. "A Transform Domain LMS Adaptive Filter with Variable Step-Size". *IEEE Signal Proc. Letters*, 9, number=2, month=Feb., year = 2002.

[10] V.J. Mathews and Z. Xie. "A Stochastic Gradient Adaptive Filter with Gradient Adaptive Step Size". *IEEE Trans. Signal Proc.*, 41(6):2075–2087, June 1993.

[11] T. Aboulnasr and K. Mayyas. "A Robust Variable Step-Size LMS-Type Algorithm: Analysis and Simulations". *IEEE Trans. on Signal Proc.*, 45(3), March 1997.

[12] W.J. Song and M.S. Park. "A Comlementary Pair LMS Algorithm for Adaptive Filtering". In *Proc. ICASSP*, Münich, Germany, Apr. 1997.

[13] W.S. Chaer, R.H. Bishop, and J. Ghosh. "A Mixture-of-Experts Framework for Adaptive Kalman Filtering". *IEEE Trans. Systems, Man and Cybernetics, Part B*, 27(3), June 1997.

[14] S.S. Kozat and A.C. Singer. "Further Results in Multistage Adaptive Filtering". In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Proc.*, March 2002.

[15] J. Arenas-Garcia, V. Gomez-Verdejo, M. Martinez-Ramon, and A.R. Figueiras-Vidal. "Separate-Variable Adaptive Combination of LMS Adaptive Filters for Plant Identification". In *Proc. IEEE NNSP Workshop*, Toulouse, France, Sept. 2003.

[16] 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.

[17] D.T.M. Slock. "Fractionally-Spaced Subband and Multiresolution Adaptive Filters". In *Proc. ICASSP*, Toronto, Canada, May 1991.

[18] R.K. Martin and Jr. C.R. Johnson. "NSLMS: a Proportional Weight Algorithm for Sparse Adaptive Filters". In *Proc. Asilomar Conf. Signals Systems and Computers*, Pacific Grove, CA, USA, 2001.

[19] J. Benesty and S.L. Gay. "An Improved PNLMS Algorithm". In *Proc. ICASSP*, Orlando, FL,USA, 2002.

[20] D. Schafhuber, G. Matz, F. Hlawatsch, and P. Loubaton. "MMSE Estimation of Time-Varying Channels for DVB-T Systems with Strong Co-Channel Interference". In *Proc. European Signal Processing Conference (EUSIPCO)*, Toulouse, France, Sept. 2002.

[21] S.S. Kozat and A.C. Singer. "Further Results in Multistage Adaptive Filtering". In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Proc.*, March 2002.

[22] 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". 45(5):1118–1128, May 1997.

[23] V. Solo and X. Kong. *"Adaptive Signal Processing Algorithms: Stability and Performance"*. Prentice-Hall, 1994.

[24] J.B. Evans, P. Xue, and B. Liu. "Analysis and Implementation of Variable Step Size Adaptive Algorithms". *IEEE Trans. Signal Proc.*, 41(8):2517–2535, Aug. 1993.

[25] Tayeb Sadiki and Dirk T. M. Slock. "Bayesian Adaptive Filtering At linear Cost". *IEEE Workshop Statistical signal Processing,Bordeaux*, Jul. 2005.