# STEADY-STATE PERFORMANCE ANALYSIS OF 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

Adaptive filtering is in principle intended for tracking nonstationary systems. However, most adaptive filtering algorithms have been designed for converging to a fixed unknown filter. When actually confronted with a non-stationary environment, they possess just one parameter (stepsize, window size) to adjust their tracking capability. In the stationary case of non-stationarity, the optimal filter coefficients evolve as a stationary process. The Bayesian approach to adaptive filtering exploits the a priori information in this stationary parameter variation model to optimize adaptive filtering performance. The prior information contains two critical parameter characteristics: the variance (magnitude) of the various filter coefficients and their variation spectrum (power delay profile and Doppler spectrum in the case of wireless channel tracking). The practical tool for implementing Bayesian Adaptive Filtering (BAF) is the Kalman filter, which typically models the parameter variation as an AR(1) process. To further limit the complexity to the same order as the complexity of the RLS algorithm, a diagonal AR(1) model can be taken. In this paper, we analyze the effect of power delay profile and Doppler bandwidth on the steady-state performance of BAF and LMS and RLS algorithms. The approximation effects of using a simplified state model are also exhibited.

## 1. INTRODUCTION

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 the initial convergence. The Recursive Least-Squares (RLS) algorithm was also developed in the 1960's and 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 predetermined. 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. This motivated on the one hand the development of fast RLS algorithms, and on the other hand the development of an intermediate category of algorithms, all less sensitive than LMS to the input correlation structure, including frequency or other transform domain LMS algorithms, prewhitened LMS versions, Fast Newton Transversal Filters and (Fast) Affine Projection Algorithms.

At the outset, all these algorithms are developed to converge to an unknown optimal filter. When this optimal filter is actual time-varying, these algorithms need to be made adaptive. The RLS algorithms are made adaptive by the introduction of a weighting function/window. The weighted LS cost function can be viewed as the output of a filter with the instantaneous squared filtering error sequence as input. The filter should be such that its input-output relationship is simple and recursive. The LS cost function uses a discrete-time integrator as filter, which can be easily modified into a first-order recursive filter for the exponentially weighted RLS algorithm. The sliding window RLS algorithm uses a moving average filter that can also be expressed recursively. All other adaptive filtering algorithms are made adaptive by the introduction of a scalar stepsize. In fact, the time-varying stepsize sequence of stochastic gradient algorithms [1] is made time-invariant/constant to avoid convergence and permit tracking of time-varying optimal filter settings. The tracking characteristics of the LMS and RLS algorithms got analyzed only in the 1970's and 1980's, 10 to 20 years after the introduction of the algorithms, in [2] for LMS and [3] for RLS. A further inspection of these tracking characteristics revealed the surprising result that in certain cases the LMS algorithm may provide better tracking

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.

than the RLS algorithm (each with optimized stepsize or forgetting factor), see [4] for deterministic and e.g. [1] for random parameter variations. With hindsight, this is not at all surprising since LMS and RLS are just two suboptimal approaches to tracking time-varying parameters. Whereas initial convergence is about the fast reduction of the mean parameter error vector, tracking is about the optimal compromise between MSE due to estimation noise and tracking/lag noise. Some general references on the tracking behavior of adaptive filtering algorithms are [5], [6], [7], [8], [9].

### 2. PROBLEM FORMULATION

Consider the problem of estimating the desired response signal d(k) as a linear combination of the elements of X(k), the N-dimensional input vector sequence to the adaptive filter. The popular adaptive filters (LMS and RLS) updates the filter coefficients in the following manner:

$$e(k) = d(k) - X^H H(k) \tag{1}$$

and

$$H(k+1) = H(k) + KX(k)e(k)$$
 (2)

where  $K = \mu$  is the step-size parameter that controls the speed convergence as well as the steady-state and /or tracking behavior of the adaptive filter in the LMS case, and  $K = R_{XX}^{-1}$  for RLS algorithm.

For the performance analysis, we will assume that the adaptive filter structure is that of an N-point FIR filter, and the input signal X(k) is obtained as a vector formed by the most recent N samples of the input sequence x(k), i.e.,

$$X(k) = [x(k), \ x(k-1), \dots, x(k-N+1)]^T.$$
(3)

Let  $H^o(k)$  denote the optimal coefficient vector (in the minimum mean-squared estimation error sense(MMSE)) for estimating the desired response signal d(k) using X(k). We will assume that  $H^o(k)$  is time varying, and that the time variations are caused by a random disturbance of the optimal coefficient process. Thus, the behavior of the optimal coefficient process can be modeled as a low pass plus a power delay profile.

In order to make the analysis tractable, we will make use of the following assumptions and approximations:

X(k), d(k) are jointly Gaussian and zero-mean random processes. X(k) is a stationary process. Moreover, {X(k), d(k)} is uncorrelated with {Xn, d(n)} if n ≠ k. This is the commonly employed independence assumption and is seldom true in practice. However, analysis reliable design rules in the past.

• The autocorrelation matrix  $R_{XX}$  of the input vector X(k) is a diagonal matrix and is given by:

$$R_{XX} = \sigma_x^2 I. \tag{4}$$

While this is a fairly restrictive assumption, it considerably simplifies the analysis. Furthermore, the white data model is valid representation in many practical systems such as digital data transmission systems and analog systems that are sampled at Nyquist rate and adapted using discrete-time algorithms.

### 3. EXCESS MEAN SQUARE ERROR (EMSE)

Consider:

$$\begin{cases} d_k = H_k^{oT} k_k + v_k \\ y_k = H_k^T X_k \end{cases}$$
(5)

where

-  $H_k^o$  denotes the optimal Filter

-  $H_k$  represents a given adaptive Filter(RLS, LMS and VSSLMS...) The a posteriori error is given by:

$$e_k = H_k^T X_k + v_k \tag{6}$$

where  $\widetilde{H}_k = H_k^o - H$ .

The estimation of the system parameters is performed by minimizing the cost function

$$J_k = F(q)e_k^2 \tag{7}$$

The Mean Squared Error can be written as:

$$MSE = E\left[e_k^2\right] = \sigma_v^2 + E\left[X_k^H \widetilde{H}_k^* \widetilde{H}_k^T X_k\right]$$
(8)

Thus, the Excess Mean Squared Error becomes:

$$EMSE = E\left[e_k^2\right] - MMSE = E\left[X_k^H \widetilde{H}_k^* \widetilde{H}_k^T X_k\right]$$
(9)

#### 4. MODELING OF STANDARD ADAPTIVE FILTERING BEHAVIOR

The adaptive filter is  $H_k$  and the a priori error  $e_k = d_k - X_k H_{k-1}$ . Consider the (complex) LMS algorithm first

$$\begin{aligned}
H_k^{lms} &= H_{k-1}^{lms} + \mu X_k^H e_k \\
&= (I - \mu X_k^H X_k) H_{k-1}^{lms} + \mu X_k^H v_k + \mu X_k^H X_k \\
&= (I - \mu X_k^H X_k) H_{k-1}^{lms} + \mu X_k^H v_k + \mu R H_k^o \\
&+ \mu (R - X_k^H X_k) (H_{k-1}^{lms} - H_k^o) \end{aligned} (10)$$

Then, assuming the adaptation speed is not too fast, we get approximately

$$H_k^{lms} = [I - (I - \mu R)q^{-1}]^{-1}\mu R(H_k^o + R^{-1}X_k^H v_k)$$
(11)

whereas the RLS filter update is of the form

$$H_k^{rls} = H_{k-1}^{rls} + \hat{R}^{-1} X_k^H e_k$$
  
=  $\frac{1-\lambda}{1-\lambda q^{-1}} (H_k^o + R^{-1} X_k^H v_k)$  (12)

In general

$$\hat{H}_k = F_{lms,rls}(q)(H_k^o + R^{-1}X_k^H v_k) = F(q)G_k.$$
 (13)

Gradient  $G_k = R^{-1}X_k^H y_k$  in fact! Then we can estimate  $S_{GG}$  assume RLS or LMS with white. where  $q^{-1}H_k = H_{k-1}$ . Using averaging analysis at low adaptation speed, these results for the sysid-up hold approximately also for the other adaptive filtering applications. Note that  $(H_k^o + R^{-1}X_k^H v_k)$  is closely related to  $G_k = R^{-1}X_k^H y_k$ , which is a mixed quantity in that it is averaged in the input covariance but instantaneous in the input-desired-response correlation.

## 5. BAYESIAN ADAPTIVE FILTERING (BAF)

In this paper we focus on stationary time-varying parameters, we neglect transient phenomena, and we consider the stationary steady-state regime. Hence it is more practical to formulate the parameter tracking problem as a Wiener filtering problem rather than a Kalman filtering problem. The kF'ing approach may be practical for an AR(1) model for the optimal parameters but becomes cumbersome for higher-order models.

We shall introduce, mostly for the purpose of analysis, a somewhat idealized Bayesian solution which is based on the assumption that R can be estimated well. this solution will be based on LMMSE estimation (WF'ing) of  $H_k^o$  from the gradient:

$$G_{k} = R^{-1}X_{k}^{H}y_{k} = H_{k}^{0} + \underbrace{R^{-1}X_{k}^{H}v_{k}}_{\tilde{G}_{k}} + (R^{-1}X^{H}X - I) H_{k}^{0}$$

where for slow parameter variations, the last term can be neglected since it is the product of low-pass noise  $H_k^o$  with high-pass noise  $R^{-1}X_k^H X_k - I$ . The optimal BAF would be to apply the kF to (11),  $G_k = H_k^o + \tilde{G}$ , which can be considered as a measurement equation for the state  $H_k^o$ . In steady-state, the Kf converge to the WF

$$H_k^o = F(q)G_k \tag{14}$$

where in the non-causal case

$$F(q) = I - S_{\tilde{G}\tilde{G}}(q)S_{GG}^{-1}(q)$$
(15)

Neglecting the last term in(11) and assuming that  $v_k$  is white noise (hence  $\tilde{G}_k$ ), we have  $S_{\tilde{G}\tilde{G}}(q) = \sigma_v^2 R^{-1}$ . Hence the non-causal WF is fairly straightforward to find since  $S_{GG}$ can be estimated simply from the observations of  $G_k$ , though  $\sigma_v^2$  is somewhat trickier to derive from the observed MSE. For the causal case, consider  $N_k = P(q)G_k$  where P(q) is the (length) (monic) multivariate prediction error filter for the vector signal  $G_k$  and  $N_k$  is resulting white prediction error with covariance matrix  $R_{NN}$ . then the causal WF is

$$F(q) = I - S_{\tilde{G}\tilde{G}}(q)R_{NN}^{-1}P(q)$$
(16)

### 6. PERFORMANCE ANALYSIS

In this section we will comparing the resulting EMSE with optimized individual stepsize to the classical LMS with an optimized global stepsize, the classical RLS with optimized individual forgetting factor, and the optimal solution (WF). The EMSE is defined in (9). If we make the assumption that the system variation is a zero-mean, wide-sense stationary process with a power spectral density matrix  $S_{HH}(e^{-2j\pi f})$ , and if we suppose that these variations are independent from the input signal, the Excess MSE can be expressed in the following form:

$$EMSE = tr\{E(\tilde{H}_{k}\tilde{H}_{k}^{H}R)\} = tr\{R\int_{-\frac{1}{2}}^{\frac{1}{2}}S_{\tilde{H}\tilde{H}^{H}}(e^{j2\pi f})df\}$$
  
$$\tilde{H}_{k} = H_{k}^{o} - H_{k} = (I - F(q))H_{k}^{o} - R^{-1}F(q)X_{k}^{*}v_{k}$$
  
$$= (I - F(q))H_{k}^{o} - F(q)\tilde{G}_{k}$$

and becomes:

$$EMSE = trR\{\int_{-\frac{1}{2}}^{\frac{1}{2}} F(e^{j2\pi f})R_{\tilde{G}\tilde{G}} F^{H}(e^{j2\pi f})df\} + trR\{\int_{-\frac{1}{2}}^{\frac{1}{2}} (I - F(e^{j2\pi f})) \times S_{H^{o}H^{o}} (e^{-j2\pi f})(I - F^{H}e^{j2\pi f})\}df$$

Remark that the EMSE can be broken up into two terms:

- $E_{noise} = trR\{\int_{-\frac{1}{2}}^{\frac{1}{2}} F(e^{j2\pi f})R_{\tilde{G}\tilde{G}} F^{H}(e^{j2\pi f})df\}$ characterizing the noise contribution; and can be interpreted as the estimation accuracy under stationary conditions
- $E_{lag} = trR\{\int_{-\frac{1}{2}}^{\frac{1}{2}} (I F(e^{j2\pi f})) S_{H^{o}H^{o}}(e^{j2\pi f}))(I F^{H}(e^{j2\pi f}))\}df$  representing the estimation error resulting from the system variations (Lag noise)

The analysis is easiest when the input is white,  $R = \sigma_y^2 I$ . We consider the uniform dynamics plus power delay profile like structured model for the optimal Doppler spectrum:  $S_{H^0H^0}(z) = S_{hh}(z) D$  where  $S_{hh}(z)$  is scalar and D is a constant diagonal. To simplify, we suppose, also, that the scalar spectrum  $S_{hh}$  is a flat low-pass spectrum; i.e.

$$S_{hh}(e^{-2j\pi f}) = \begin{cases} 1 & \text{if } f < f_0\\ 0 & \text{elsewhere} \end{cases}$$

The matrix D is arbitrary if it were diagonal (decorrelation filter coefficients) the diagonal would represent the power delay profile of the optimal filter

We deduce, thus, the EMSE expressions for the different cases:

in the RLS case

$$F(z) = \frac{1-\lambda}{1-\lambda z^{-1}}I,$$

$$EMSE^{RLS} = N\sigma_v^2 \frac{1-\lambda}{1+\lambda} + 2\sigma_x^2 tr(D)$$

$$\times (\lambda f_o - \frac{\lambda}{\sigma} \frac{1-\lambda}{1+\lambda} \arctan(\frac{1+\lambda}{1+\lambda} \tan(\pi f_o)))$$
(17)

in the LMS case

$$F(z) = \frac{\mu R}{I - (I - \mu R)z^{-1}}$$
(18)

$$\begin{split} EMSE^{LMS} &= N\sigma_v^2 \frac{\mu \sigma_x^2}{2-\mu \sigma_x^2} + 2\sigma_x^2 (1-\mu \sigma_x^2) tr(D) \\ &\times \left(f - \frac{\mu \sigma_x^2}{\pi (2-\mu \sigma_x^2)} (\arctan \frac{\mu \sigma_x^2}{2-\mu \sigma_x^2} \tan(\pi f))\right) \end{split}$$

### in the LMS with Individual StepSize case (LMSISS)

The LMSISS type adaptive algorithm is a gradient search algorithm which computes a set of weights  $H_k$  that seeks to minimize  $E(d_k - Y_k^T H_k)^2$ . The algorithm is of the form:

$$\hat{H}_k = \hat{H}_{k-1} + M_k Y_k e_k , \quad e_k = d_k - H_k^T Y_k$$

and  $M_k$  is a diagonal matrix containing the different variable step sizes. In the standard LMS algorithm,  $M_k = \mu_k I$  with  $\mu_k$  a constant.

In the non-causal case

$$F(q) = I - S_{\tilde{G}\tilde{G}}(q)S_{GG}^{-1}(q)$$
(19)

the EMSE expressions being:

$$EMSE_{ncc} = \sum_{i=1}^{N} \frac{1}{j2\pi} \oint \frac{dz}{z} (\frac{1}{\sigma_v^2} + \frac{1}{\sigma_x^2} D_{ii} S_{hh}(z))^{-1}$$
$$= \sigma_v^2 2f_o \sum_{i=1}^{N} \frac{1}{1 + \frac{\sigma_v^2}{\sigma_x^2} D_{ii} \frac{1}{2f_o}}$$

we compare the minimum EMSE achieved by each variant (with optimized parameters  $\lambda; \mu$ ). Figure 1 plots the optimized EMSE curves (as a function of the Doppler bandwidth  $f_0$ ).

# 7. CONCLUSION

In this paper we focus on stationary time-varying parameters, we neglect transient phenomena, and we consider the stationary steady-state regime. Hence it is more practical to formulate the parameter tracking problem as a Wiener filtering problem rather than a Kalman filtering problem. This analysis shows that, for a flat low-pass spectrum, the optimal RLS performs better that the optimal LMS and LM-SIVSS, but also that the Baysian Adaptive Feltering (BAF) given here with a causal and non-causal Wiener filter performs even better.



**Fig. 1**. *EMSE*<sub>opt</sub> curves for ideal low-pass variations.

### 8. REFERENCES

- D.T.M. Slock, ""On the Convergence Behavior of the LMS and the Normalized LMS Algorithms"," *IEEE Trans. on Signal Processing*, vol. 41, no. 9, pp. 2811– 2825, Sept. 1993.
- [2] B. Widrow *et al.*, ""Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter"," *Proc. IEEE*, vol. 64, no. 8, pp. 1151–1162, Aug. 1976.
- [3] L. Ljung and S. Gunnarsson, ""Adaptive Tracking in System Identification, A Survey"," *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.
- [4] O.M. Macchi and N.J. Bershad, ""Adaptive Recovery of a Chirped Sinusoid in Noise, Part I: Performance of the RLS Algorithm, Part II: Performance of the LMS Algorithm"," *IEEE Trans. SP*, vol. SP-39, pp. 583–602, March 1991.
- [5] Tayeb Sadiki and Dirk T. M. Slock, "Bayesian Adaptive Filtering At linear Cost"," *IEEE Workshop Statisti*cal signal Processing, Bordeaux, Jul. 2005.
- [6] L. Ljung and T. Soderstrom, "Theory and Practice of Recursive Identification", MIT Press, Cambrdge, MA, 1983.
- [7] V. Solo and X. Kong, "Adaptive Signal Processing Algorithms: Stability and Performance", Prentice-Hall, 1994.
- [8] M. Niedzwiecki, "Identification of Time-Varying Systems", Wiley, 2000.
- [9] B.D.O. Anderson and J.B. Moore, "Optimal Filtering", Prentice-Hall, Englewood Cliffs, NJ, 1979.