

FURTHER RESULTS IN MULTISTAGE ADAPTIVE FILTERING

Suleyman S. Kozat and Andrew C. Singer

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801 USA
Email: {kozat, singer}@ifp.uiuc.edu

ABSTRACT

In this paper, we investigate some of the stochastic properties of two recently introduced multistage adaptive filtering algorithms, namely the LMS-Bayesian and the RLS-Bayesian algorithms. We study probability-1 convergence of these algorithms and derive their final mean squared error for stationary Gaussian time series. We will show that under some general independence assumptions, both algorithms are convergent in a probability-1 sense and achieve the performance of the best algorithm used in the mixture.

1. INTRODUCTION

The performance of multistage adaptive filtering (MSAF) in general is studied in a number of recent papers [1][4]. In these papers, the structure of MSAF was introduced, however the theoretical derivations were mainly focused on mean square error curves and the assumptions invoked were rather strong. In this paper, we focus on the convergence behavior of these algorithms and investigate their final mean squared error.

We define a multistage adaptive filter as one whose output is obtained by combining the outputs of multiple constituent adaptive filters, linear or nonlinear, with the goal of outperforming the best algorithm among them. The first stage of a multistage adaptation algorithm consists of m constituent algorithms that operate in parallel on the observation sequence. The j th algorithm outputs $\hat{d}_j(n)$, which is compared with the observed (desired) data $d(n)$, and the error $e_j(n)$ is fed back to the adaptation algorithm. Each algorithm operates in parallel, with its own adaptation process, where the outputs of the algorithms are decoupled from one another. Depending on the application, these constituent algorithms may include a wide variety of models.

The second stage of the multistage algorithm is the model mixture stage, where the final output, $\hat{d}_u(n)$, and the adaptation process is given by

$$\begin{aligned}\hat{d}_u(n) &= \sum_{k=1}^m u_k(n) \hat{d}_k(n), \\ e_u(n) &= d(n) - \hat{d}_u(n), \\ \underline{u}(n+1) &= \underline{f}_u(e_u(n), \underline{u}(n), \hat{\underline{d}}_u(n)),\end{aligned}\quad (1)$$

where \underline{f}_u is the vector adaptation function, $\hat{\underline{d}}_u(n) = [\hat{d}_1(n), \dots, \hat{d}_m(n)]^T$, and $\underline{u}(n) = [u_1(n), \dots, u_m(n)]^T$. In this mixture stage, the outputs of the first stage algorithms are

adaptively combined to give the final output. Depending on the application, the first stage outputs can be combined by explicitly taking into account their performance on the observed data so far, or implicitly through another adaptation algorithm operating on the outputs of the first stage.

In this paper, we will investigate two algorithms specifically for the problem of prediction: the LMS-Bayesian and RLS-Bayesian algorithms. In Section 2, we define LMS-Bayesian algorithm and investigate its convergence behavior. We will show that with basic independence assumptions the LMS-Bayesian algorithm is convergent in a probability-1 sense and achieves the performance of the best algorithm used in the mixture. In Section 3, after introducing the RLS-Bayesian algorithm, we carry out a similar line of analysis. With some basic assumptions we will show that the RLS-Bayesian algorithm is also convergent in a probability-1 sense and achieves the performance of the best linear predictor for each sequence.

2. LMS-BAYESIAN ALGORITHM

Let $\hat{x}_k(n)$ be the output of a sequential linear predictor as obtained by the LMS algorithm with model order k , i.e.,

$$\begin{aligned}\hat{x}_k(n) &= \underline{w}_k^T(n-1) \underline{x}_k(n-1), \\ e_k(n) &= x(n) - \hat{x}_k(n), \\ \underline{w}_k(n) &= \underline{w}_k(n-1) + \mu e_k(n) \underline{x}_k(n-1),\end{aligned}\quad (2)$$

where μ is a constant to control stability and rate of convergence, and $e_k(n)$ is the prediction error to be minimized in the mean-square. The weight and input vectors are given by $\underline{w}_k(n-1) = [w_{1,k}^{n-1}, \dots, w_{k,k}^{n-1}]^T$ and $\underline{x}_k(n) = [x(n), \dots, x(n-k+1)]^T$, respectively. Define the LMS-Bayesian linear predictor as the following weighted sum over linear predictors of order less than or equal to m :

$$\begin{aligned}\hat{x}_{LB}(n) &= \sum_{k=1}^m u_k(n) \hat{x}_k(n), \\ u_k(n) &= \frac{\exp(-c l_{t-1}(x, \hat{x}_k))}{\sum_{k=1}^m \exp(-c l_{t-1}(x, \hat{x}_k))},\end{aligned}\quad (3)$$

where c is a positive constant and $u_k(n)$, the weights in the mixture, are proportional to the performance of the k th-order predictor on the data observed so far. The performance, $l_{t-1}(x, \hat{x}_k) = \sum_{i=1}^{t-1} (x(i) - \hat{x}_k(i))^2$, is the accumulated squared prediction error that results from $\hat{x}_k(t)$.

By using two independence assumptions, Assumption 1 and Assumption 2 below, it can be shown that:

Statement 1: For a wide-sense stationary Gaussian observation process, the mixture coefficients of the LMS-Bayesian algorithm $u_k(n)$ are consistent in probability-1, such that

$$\begin{aligned} u_k(n) &\rightarrow 1, \text{ as } n \rightarrow \infty \text{ (pr)}, k = \min(w, m), \\ u_k(n) &\rightarrow 0, \text{ as } n \rightarrow \infty \text{ (pr)}, k \neq \min(w, m), \end{aligned} \quad (4)$$

where w is the order of the MMSE-optimal linear filter of order less than or equal to m . For example, if the underlying process, $x(n)$ is given by $x(n) = \sum_{k=1}^w c_k x(n-k) + \varepsilon(n)$, where $\varepsilon(n)$ is a sequence of i.i.d. Gaussian random variables with zero mean and variance σ_ε^2 , then w is the order of this auto-regressive process. For a general stationary Gaussian process, the required order can be arbitrarily large, thus all the coefficients except $k = m$ will vanish. The independence assumptions we invoke are:

Assumption 1: At any time n , the observation sequence and the input vectors to the predictors are independent of the past samples of the observation sequence and the input vectors. This is the independence assumption from the adaptive signal processing literature usually used for the LMS algorithm [3].

Assumption 2: $\text{Var}(Y_{p,k}(n)) = o(n^2)$, where $Y_{p,k}(n) \triangleq \sum_{l=1}^n (e_p^2(l) - e_k^2(l))$.

For any p , from the definition in Equation (3), the coefficients $u_k(n)$ can be expressed

$$u_p(n) = \frac{1}{\sum_{k=1}^m \exp(-c Y_{p,k}(n))}, \quad (5)$$

where for any p and k , $Y_{p,k}(n) = \sum_{l=1}^n (e_p^2(l) - e_k^2(l))$. By the Chebyshev inequality, for any $\epsilon \in R^+$,

$$\text{Pr} \left[\left| \frac{Y_{p,k}(n) - E[Y_{p,k}(n)]}{n} \right| < \epsilon \right] > 1 - \frac{\text{Var}(Y_{p,k}(n))}{n^2 \epsilon^2}. \quad (6)$$

From Assumption 2, $\text{Var}(Y_{p,k}(n)) = o(n^2)$, we conclude that

$$\frac{\text{Var}(Y_{p,k}(n))}{n^2} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

For calculating the term $E[Y_{p,k}(n)]$ in Equation (6), we rely upon the widely used set of ‘‘independence assumptions’’ [3] detailed in Assumption 1, often used to derive the expected squared prediction error of an LMS linear predictor. For an LMS predictor of order p , the prediction error at any time n satisfies

$$E[e_p^2(n)] = J_{\min,p} \left(1 + \sum_{k=1}^p \frac{\lambda_k \mu}{1 - 2\mu\lambda_k} \right) + \sum_{k=1}^p \beta_{k,p} \alpha_{k,p}^n, \quad (7)$$

where λ_k is the k th eigenvalue of the correlation matrix $E[\underline{x}_k(n)\underline{x}_k^T(n)]$, and $J_{\min,p}$ exists and is the expected square error for the optimal p th-order predictor. The geometric terms $\alpha_{k,p}^n$ are the eigenvalues of a certain $p \times p$ matrix, where $|\alpha_k| \leq 1$ if the first stage algorithms are convergent [3]. The quantity $J_{\min,p}$ is a nonincreasing function of p such that the optimal p th-order linear predictor asymptotically outperforms (or at least gives the same minimum error of) any predictor with order less than p . For

small values of μ , Equation (7) can be approximated such that $1 + \sum_{k=1}^p \frac{\lambda_k \mu}{1 - 2\mu\lambda_k} \approx 1 + \sum_{k=1}^p \lambda_k \mu$, which is equal to $(1 + \mu p \sigma_x^2)$. Therefore, $E[Y_{p,k}(n)]$ will be $E[Y_{p,k}(n)] \approx J_{\min,p}(1 + \mu p \sigma_x^2)n - J_{\min,k}(1 + \mu k \sigma_x^2)n + g(n)$, where

$g(n) = \sum_{l=1}^p \beta_{l,p} \alpha_{l,p}^n - \sum_{l=1}^k \beta_{l,k} \alpha_{l,k}^n$. Since the geometric terms $g(n)$ are $o(n)$, their contributions will be negligible in $E[Y_{p,k}(n)]$ compared to the terms linear in n . Suppose the underlying process to be estimated $x(n)$ is a w th-order Gaussian AR process, with innovation sequence $\varepsilon(n)$ with variance σ_ε^2 . When $(m > w)$, the term $J_{\min,p}$ is a monotonically nonincreasing function of p . For sufficient order predictors ($p \geq w$), $J_{\min,p} = J_{\min,w} = \sigma_\varepsilon^2$. Then, for any predictor with order ($p > w$),

$$\begin{aligned} E[Y_{p,w}(n)] &= c \left(J_{\min,w}(1 + \mu p \sigma_x^2) - J_{\min,w}(1 + \mu w \sigma_x^2) \right) n + o(n), \\ &= A n + o(n), \end{aligned} \quad (8)$$

where $A \in R^+$. By Equation (6) this yields,

$$\text{Pr} [|Y_{p,w}(n) - An| < n\epsilon] > 1 - \frac{\text{var}(Y_{p,w}(n))}{n^2 \epsilon^2},$$

$$\text{Pr} [(A - \epsilon)n < Y_{p,w}(n)] \geq$$

$$\text{Pr} [|Y_{p,w}(n) - An| < n\epsilon] \geq 1 - \frac{\text{var}(Y_{p,w}(n))}{n^2 \epsilon^2}.$$

Selecting $\epsilon < A$, this inequality leads to $Y_{p,w}(n) \rightarrow \infty$, as $n \rightarrow \infty$ (pr), which implies $\exp(c Y_{p,w}) \rightarrow \infty$, as $n \rightarrow \infty$ (pr). This last limit implies that for sufficient order predictors, at least one of the terms in the denominator of Equation (5) will diverge. Thus for ($p > w$), $u_p(n) \rightarrow 0$, as $n \rightarrow \infty$ (pr). For any insufficient order predictor with order ($p < w$),

$$\begin{aligned} E[Y_{p,w}(n)] &= \\ &= c \left(J_{\min,p}(1 + \mu p \sigma_x^2) - J_{\min,w}(1 + \mu w \sigma_x^2) \right) n. \end{aligned} \quad (9)$$

Unlike Equation (8), in Equation (9) the sign of the difference term is not obvious. Although, $J_{\min,p}$ strictly decreases in p for ($p < w$), the linear term $\mu p \sigma_x^2$ increases in p .

The difference term in Equation (9) must be positive to guarantee that the weights of the insufficient order terms in the mixture go to zero in probability as n increases. Thus, certain conditions must be imposed on μ , to balance these counteracting terms to give a positive difference. We observe that $E[Y_{p,w}(n)] = Bn + o(n)$, where $B \in R^+$, when $J_{\min,w}(1 + \mu w \sigma_x^2) < J_{\min,p}(1 + \mu p \sigma_x^2)$ for any $p < w$. With this condition at least one term will diverge in the denominator of $u_p(n)$ in Equation (5). For this condition, we must have

$$\frac{(1 + \mu p \sigma_x^2)}{(1 + \mu w \sigma_x^2)} > \frac{\sigma_\varepsilon^2}{\sigma_x^2 - \underline{p}_p^T R_k \underline{p}_p} \triangleq a, \quad (10)$$

where $\underline{p}_p = E[x(n)\underline{x}_p(n)]$ is the cross correlation vector, and $R_p = E[\underline{x}_p(n)\underline{x}_p^T(n)]$ is the correlation matrix of the input, which is assumed to be positive definite. Because $J_{\min,p}$ is monotonically decreasing in p for $p \leq w$ [3], the ratio a on the right-hand side of Equation (10) is always

between 0 and 1. Since, the left hand side of Equation (10) is also monotonically decreasing from 1 to p/w , as μ increases from 0 to ∞ , there will always be a nontrivial interval such that the condition given in Equation (10) is satisfied. The interval is given by

$$\frac{1-a}{w\sigma_x^2 a - p\sigma_x^2} > \mu > 0, \quad (11)$$

which provides from Equation (5) $u_p(n) \rightarrow 0$, as $n \rightarrow \infty$ (pr), where ($p < w$). Since, $\sum_{k=1}^m u_k(n) = 1$, we conclude that $u_w(n) \rightarrow 1$, as $n \rightarrow \infty$ (pr), which is convergence in probability of the weights $u_k(n)$. Moreover, by definition $0 < u_k(n) < 1$ for all n and k , which implies $\text{Var}(u_k(n)) < 1$. These bounded variances, together with convergence in probability yield,

$$E[u_k(n)] \rightarrow 1, \text{ as } n \rightarrow \infty, k = \min(w, m)$$

$$E[u_k(n)] \rightarrow 0, \text{ as } n \rightarrow \infty, \text{ else,} \quad (12)$$

with the same notation as in Equation (4).

Assumption 3: The weights of predictors are independent from $x(n)$ and $\hat{x}_k(n)$.

Based on Assumption 3, we can make the following statement.

Statement 2:

$$\lim_{n \rightarrow \infty} E[(x(n) - \hat{x}_{LB}(n))^2] \leq \lim_{n \rightarrow \infty} \min_{k=1, \dots, m} E[(x(n) - \hat{x}_k(n))^2],$$

which implies that the LMS-Bayesian algorithm is asymptotically better than the best of the mixture algorithms in terms of the final MSE.

To derive this result, observe that the MSE of the LMS-Bayesian is given by

$$E[(x(n) - \hat{x}_{LB}(n))^2] = E\left[\left(\sum_{k=1}^m u_k(n)(x(n) - \hat{x}_k(n))\right)^2\right], \quad (13)$$

since $\sum_{k=1}^m u_k(n) = 1$. By the convexity of the square function,

$$E[(x(n) - \hat{x}_{LB}(n))^2] \leq E\left[\sum_{k=1}^m u_k(n)(x(n) - \hat{x}_k(n))^2\right].$$

With Assumption 3 that the weights of predictors are independent from $x(n)$ and $\hat{x}_k(n)$ at time n , we can conclude

$$E[(x(n) - \hat{x}_{LB}(n))^2] \leq \sum_{k=1}^m E[u_k(n)]E[(x(n) - \hat{x}_k(n))^2],$$

which implies, by Equation (12), that

$$\lim_{n \rightarrow \infty} E[(x(n) - \hat{x}_{LB}(n))^2] \leq \lim_{n \rightarrow \infty} \min_{k=1, \dots, m} E[(x(n) - \hat{x}_k(n))^2]. \quad (14)$$

3. RLS-BAYESIAN ALGORITHM (ULP)

The linear predictors used in first stage of the LMS-Bayesian algorithm, defined in Section 2, update their weights using the LMS algorithm. In this section, the coefficients will be updated using the RLS algorithm. Let $\hat{x}_k(n)$ be the output of a sequential linear predictor as obtained by an RLS algorithm with model order k , i.e.,

$$\begin{aligned} \hat{x}_k(n) &= \underline{w}_k^T(n-1)\underline{x}_k(n-1), \\ e_k(n) &= x(n) - \hat{x}_k(n), \end{aligned}$$

where at any time n , the predictor coefficients $\underline{w}_k(n-1) = [w_{1,k}^{n-1}, \dots, w_{k,k}^{n-1}]^T$ are obtained such that the total squared prediction error over the past,

$$E_{n-1}(x, \hat{x}_k) = \sum_{t=1}^{n-1} \left(x(t) - \sum_{i=1}^k w_{i,k} x(t-i) \right)^2, \quad (15)$$

is minimized over these coefficients. Define a new predictor $\hat{x}_{RB}(n)$ as a weighted sum over the linear predictors of order less than or equal to m , as in Equation (3).

Let $Y_{p,k}(n) \triangleq \sum_{l=1}^n (e_p^2(l) - e_k^2(l))$. Also, we make the following assumption.

Assumption 4: $\text{Var}(Y_{p,k}(n)) = o(n^2)$, where $Y_{p,k}(n) \triangleq \sum_{l=1}^n (e_p^2(l) - e_k^2(l))$.

As a result, we can show the following.

Statement 3: For a wide sense stationary Gaussian observation process, the mixture coefficients of the RLS-Bayesian algorithm, $u_k(n)$, for insufficient order predictors vanish with probability-1 as $n \rightarrow \infty$, such that,

$$u_k(n) \rightarrow 0, \text{ as } n \rightarrow \infty, k < \min(w, m) (pr) \quad (16)$$

where w is the smallest order such that

$E[(x(n) - \hat{x}_w(n))^2] = E[(x(n) - \hat{x}_m(n))^2]$. By the definition of $Y_{p,k}(n)$ and using the Chebyshev inequality, for any $\epsilon \in R^+$,

$$Pr\left[\left|\frac{Y_{p,k}(n) - E[Y_{p,k}(n)]}{n}\right| < \epsilon\right] > 1 - \frac{\text{Var}(Y_{p,k}(n))}{n^2\epsilon^2}. \quad (17)$$

As in Section 3.1.1, we will find an expression for $E[Y_{p,k}(n)]$ in Equation (17) and bound the $Y_{p,k}(n)$ term.

In this probabilistic setting, Davisson has shown [2] that the expected squared prediction error of an RLS linear predictor of order p for any n satisfies,

$$E[e_p^2(n)] = J_{min,p}(1 + \frac{p}{n}) + o(1/n). \quad (18)$$

This implies, $E[Y_{p,k}(n)] = J_{min,p}(n + p \ln(n)) - J_{min,k}(n + k \ln(n)) + o(\ln(n))$. If the underlying process is $x(n)$ is a w th-order Gaussian AR process, then for any predictor of order ($p < w$)

$$E[Y_{p,w}(n)] = An + O(\ln(n)), \quad (19)$$

where $A \in R^+$. By Equation (17), this yields

$$\begin{aligned} Pr[(A - \epsilon)n < Y_{p,w}(n)] &\geq \\ Pr[|Y_{p,w}(n) - An| < n\epsilon] &\geq 1 - \frac{\text{Var}(Y_{p,w}(n))}{n^2\epsilon^2}. \end{aligned}$$

Selecting $\epsilon < A$, this inequality leads to $Y_{p,w}(n) \rightarrow \infty$, as $n \rightarrow \infty$ (pr), which implies $\exp(cY_{p,w}) \rightarrow \infty$, as $n \rightarrow \infty$ (pr). This last limit implies that for insufficient order predictors, at least one of the terms in the denominator of Equation (5) will diverge. Thus for $(p < w)$, $u_p(n) \rightarrow 0$, as $n \rightarrow \infty$ (pr). With the same reasoning as in Section 2, it can be shown that the mean of the weights, $u_k(n)$, converge to the same value as n goes to infinity, i.e.,

$$E[u_p(n)] \rightarrow 0, \text{ as } n \rightarrow \infty \text{ (pr)}. \quad (20)$$

Again by Assumption 3, we have the following.

Statement 4: The RLS-Bayesian algorithm is universal in the mean-square error sense

$$\lim_{n \rightarrow \infty} E[(x(n) - \hat{x}_{RB}(n))^2] \leq \min_{\hat{b} \in R^m} E[(x(n) - \hat{b}^T \underline{x}_m(n-1))^2],$$

such that it achieves the performance of the best linear predictor of order up to m .

The derivation of this result exactly follows the Statement 1 in Section 2 until Equation (14). Since from Equation (18), for all sufficient order predictors ($p \geq w$),

$$\lim_{n \rightarrow \infty} E[(x(n) - \hat{x}_p(n))^2] = \min_{\hat{b} \in R^m} E[(x(n) - \hat{b}^T \underline{x}_m(n-1))^2],$$

we conclude that by Equation (20),

$$\lim_{n \rightarrow \infty} E[(x(n) - \hat{x}_{RB}(n))^2] \leq \min_{\hat{b} \in R^m} E[(x(n) - \hat{b}^T \underline{x}_m(n))^2].$$

For convergence of the weight coefficients of the strictly sufficient order predictors ($p > w$), the previous argument is not valid. For sufficient order predictors $p \geq w$, we require the following assumptions:

Assumption 5: Each difference in the sum in Equation (23) ($e_p(l)^2 - e_w(l)^2$) is independent from $(e_p(s)^2 - e_w(s)^2)$ for $s \neq l$, the errors $e_p(n)$ are Gaussian distributed random variables, and $E[e_p(n)e_w(n)] = J_{min,w} + O(1/n)$.

As a result, we can show the following.

Statement 5: For a wide sense stationary Gaussian observation process, the mixture coefficients of the RLS-Bayesian algorithm, $u_k(n)$, for strictly sufficient order predictors vanish with probability-1 as $n \rightarrow \infty$, such that,

$$u_k(n) \rightarrow 0, \text{ as } n \rightarrow \infty, k > \min(w, m) \text{ (pr)} \quad (21)$$

The Chebyshev inequality in Equation (17) becomes $Pr[B \ln n - \epsilon < Y(n)] \geq 1 - (\text{Var}(Y(n))/\epsilon^2)$, where $B = (J_{min,w}(p-w))$, and the $o(\ln n)$ terms are negligible compared to terms in $\ln n$. If we chose $\epsilon = B_1 \ln n$ where $0 < B_1 < B$, this yields

$$Pr[(B - B_1) \ln n < Y(n)] \geq 1 - \frac{\text{Var}(Y(n))}{B_1^2 \ln^2 n}. \quad (22)$$

Thus if we can show that $(\text{Var}(Y(n))/(B_1^2 \ln^2 n)) \rightarrow 0$, then we can conclude that $Y(n) \rightarrow \infty$ as $n \rightarrow \infty$ in probability.

By definition $\text{Var}(Y(n))$ is given by,

$$\text{Var}(Y(n)) = \text{Var} \left(\sum_{l=1}^n (e_p(l)^2 - e_w(l)^2) \right). \quad (23)$$

By Assumption 5, we can interchange the variance and summation, and Equation (23) becomes $\text{Var}(Y(n)) \approx$

$\sum_{l=1}^n \text{Var}(e_p(l)^2 - e_w(l)^2)$. For each term in the above summation,

$$\begin{aligned} \text{Var}(e_p(l)^2 - e_w(l)^2) &= \text{Var}(e_p(l)^2) \\ &+ \text{Var}(e_w(l)^2) - 2(E[e_p(l)^2 e_w(l)^2] - \hat{e}_p(l)^2 \hat{e}_w(l)^2), \end{aligned} \quad (24)$$

where $\hat{e}_p(l)^2 \triangleq E[e_p(l)^2]$. By Assumption 5 and Gaussian moment factorization, it can be shown that $\text{Var}(e_p(l)^2) = 2E[e_p(l)^2]^2$, and $E[e_p(l)^2 e_w(l)^2] = \hat{e}_p(l)^2 \hat{e}_w(l)^2 + 2E[e_p(l)e_w(l)]^2$. Thus Equation (24) becomes

$$\begin{aligned} \text{Var}(e_p(l)^2 - e_w(l)^2) &= \\ &2\hat{e}_p(l)^2 + 2\hat{e}_w(l)^2 - 4E[e_p(l)e_w(l)]^2. \end{aligned} \quad (25)$$

Using Equation (18), Equation (25) yields,

$$\begin{aligned} \text{Var}(e_p(l)^2 - e_w(l)^2) &= \\ &2J_{min,w} + 2\frac{p+k}{l} - 4E[e_p(l)e_w(l)]^2 + o(1/l). \end{aligned} \quad (26)$$

By Assumption 5, $E[e_p(l)e_w(l)] = J_{min,w} + O(1/l)$, and we conclude that $\text{Var}(Y(n)) = o(\ln^2 n)$. Hence, by Equation (22) $u_k(n) \rightarrow 0$, as $n \rightarrow \infty$, $k > \min(w, m)$ (pr).

4. CONCLUSION

In this paper, two examples of MSAF algorithms were analyzed in terms of their convergence characteristics and final MSE. The MSE of the LMS-Bayesian and RLS-Bayesian algorithms are shown to converge to the MSE of the best predictor used in the MSAF. Thus, these two algorithms are shown to be universal in this stochastic context.

5. REFERENCES

- [1] A. C. Singer, M. Feder, "Universal Linear Prediction by Model Order Weighting," *IEEE Trans. on Signal Proc.*, vol. 47, no. 10, pp. 2685-2700, Oct. 1999.
- [2] L. D. Davison, "The Prediction Error of Stationary Gaussian Time Series of Unknown Covariance," *IEEE Trans. on Information Theory*, vol. 11, no. 4, pp. 527-532, Oct. 1965.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ 07458, 1996.
- [4] S. S. Kozat, A. C. Singer, "On Universal Linear Prediction of Gaussian Data", *Proc. Int. Conf. Acous. Speech and Sig. Process, Istanbul, 2000*