Adaptive AR Modeling in White Gaussian Noise

Wen-Rong Wu and Po-Cheng Chen

***Abstract*— Autoregressive (AR) modeling is widely used in signal processing. The coefficients of an AR model can be easily obtained with a least mean square (LMS) prediction error filter. However, it is known that this filter gives a biased solution when the input signal is corrupted by white Gaussian noise. Treichler suggested the γ-LMS algorithm to remedy this problem and proved that the mean weight vector can converge to the Wiener solution. In this paper, we develop a new algorithm that extends works of Vijayan *et al.* for adaptive AR modeling in the presence of white Gaussian noise. By theoretical analysis, we show that the performance of the new algorithm is superior to the γ-LMS filter. Simulations are also provided to support our theoretical results.**

I. Introduction

THE AR modeling technique has been successfully used in wide range of applications such as speech analysis [1], [2], spectrum estimation [3], [4], and noise cancellation [5]. Given a random signal, the main task of this technique is to find the optimal AR coefficients that minimize a mean-square error (MSE) criterion. When second-order statistics of the signal are known, optimal coefficients can be obtained by solving the Wiener–Hopf equations. However, these statistics are not always available in real applications. A common alternative is to use adaptive filtering techniques.

The adaptive prediction error filter is an all-zero filter that adaptively adjusts its weights to flatten the spectrum of the output signal. Such filters start from initial conditions that contain no desired information and then update their filter weights based on a sequence of input data. For stationary inputs, it has been shown that with a proper algorithm, the mean weight vector of adaptive prediction error filters will converge to optimal AR coefficients (Wiener solutions) [6]. Commonly used adaptive algorithms include the recursive least-squares (RLS) and least-mean-square (LMS) [7]. Although the RLS algorithm may give a rapid convergence, its computational complexity is high. Thus, the LMS algorithm is preferred in many real-word applications.

In practice, the input signal often contains white Gaussian noise. Thus, the signal spectrum will become flatter than the original [8]. The LMS prediction error filter, in trying to flatten the distorted spectrum, then converges to a biased solution. To overcome this drawback, Treichler [9] suggested the γ-LMS filter and showed that the mean weight vector can converge to the Wiener solution. In this paper, we approach the problem by extending works of Vijayan *et al.* [10]–[11], in which they proposed a nonlinear prediction error filter to suppress the narrowband interference in direct-sequence (DS) spread spectrum systems. The nonlinearity of the filter is due to the non-Gaussian (binary) spread signal. Using simulation results, Vijayan *et al.* showed that the nonlinear filter is superior to a linear one. Recently, this filter was used by Kim and Efron [12] for robust impulse noise filtering. For convenience, we called the prediction error filter the ρ-LMS filter, where $\rho(\cdot)$ is the nonlinear function used in [10].

Our contribution can be divided into two parts. First, we derive the second-order statistic of the γ-LMS filter, which is often used to measure the performance of an adaptive algorithm. This result was not shown in [9]. Second, we use estimation theory to derive the ρ-LMS filter and apply it to the AR modeling problem. Specifically, we developed a linear ρ-LMS filter for signals corrupted by white Gaussian noise. We derived the first- and second-order statistics of the linear ρ-LMS filter to show that it performs better than the γ-LMS filter. This paper is organized as follows. Section II states the bias effect in the LMS prediction error filter caused by white Gaussian noise. In Section III, we describe the γ-LMS filter and derive second-order statistics. In Section IV, we develop the general ρ-LMS filter. Specifically, we focus on a linear one and derive its first- and second-order statistics. In Section V, we report simulation results and draw conclusions in Section VI.

II. The LMS Prediction Error Filter in White Noise

A signal x_k modeled as a pth-order AR process can be expressed as

$$x_k = \sum_{j=1}^{p} w_j x_{k-j} + d_k \tag{1}$$

where d_k is the prediction error, and $w_1, w_2, \ldots, w_k$ are AR coefficients. It has been shown that if p is large enough, d_k is a white sequence [6]. The main task of AR modeling is to find optimal AR coefficients that minimize the mean square value of the prediction error. Let $X_k = [x_{k-1} \;\; x_{k-2} \;\; \cdots \;\; x_{k-p}]^T$ be the input vector. The optimal coefficient vector W^* is known to be the Wiener solution given by

$$W^* = [w_1^* \quad w_2^* \quad \cdots \quad w_p^*]^T = R^{-1}P \tag{2}$$

where $R = E\{X_k X_k^T\}$ is the correlation matrix, and $P = E\{x_k X_k\}$.

Manuscript received November 11, 1995; revised December 3, 1996. This work was supported by the National Science Council, Taiwan, Republic of China, under Grant NSC 85-2213-E-009-012. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jose Carlos M. Bermudez.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: wrwu@cc.nctu.edu.tw).

Publisher Item Identifier S 1053-587X(97)03342-4.

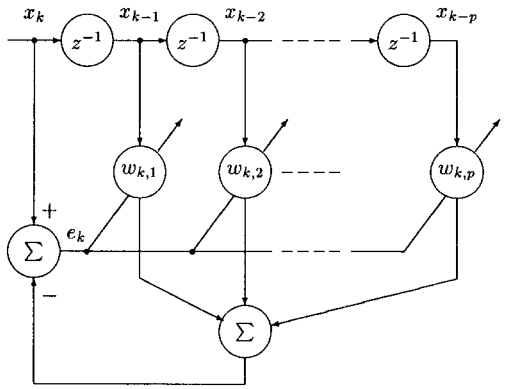

Fig. 1. LMS prediction error filter.

The LMS prediction error filter, which is illustrated in Fig. 1, can be used to adaptively estimate the optimal AR coefficients. The weight-update equation is given as follows:

$$W_{k+1} = W_k + \mu X_k e_k \tag{3}$$

where

$$e_k = x_k - \bar{x}_k \tag{4}$$

is the prediction error, and

$$\bar{x}_k = X_k^T W_k \tag{5}$$

is the prediction of x_k. The step size μ in (3) determines the rate of convergence and stability of the weights. It has been shown [7] that when μ is chosen properly, the mean weight vector will converge to the Wiener solution. If $W_k = W^*$, $\bar{x}_k$ is the optimal prediction of x_k, and the minimum mean-square error (MMSE) will be

$$J_{\min} = E\{(x_k - X_k^T W^*)^2\} = \sigma_x^2 - P^T W^* \tag{6}$$

where σ_x^2 is the variance of x_k.

Define the weight-error vector $\boldsymbol{\epsilon}_k$ and its correlation matrix K_k as

$$\boldsymbol{\epsilon}_k = W_k - W^* = [\epsilon_{k,1} \quad \epsilon_{k,2} \quad \cdots \quad \epsilon_{k,p}]^T \tag{7}$$

and

$$K_k = E\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T\}. \tag{8}$$

The excess MSE is given by

$$J_k^{\text{ex}} = E\{e_k^2\} - J_{\min} = \text{tr}(RK_k) \tag{9}$$

where $\text{tr}(\cdot)$ denotes the trace operation. To measure the performance of the LMS algorithm, the mean-square weight difference (MSD) [13], which is defined as follows, is frequently used.

$$M_k = E\left\{\sum_{i=1}^{p} \epsilon_{k,i}^2\right\} = \text{tr}(E\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T\}) = \text{tr}(K_k). \tag{10}$$

In the steady state, the correlation matrix of weight-error vector K_k can be approximated by [7]

$$K_\infty \approx \frac{1}{2}\mu J_{\min} I \tag{11}$$

where I is a $p \times p$ identity matrix. Thus, the steady-state MSD is

$$M_\infty \approx \frac{1}{2}\mu p J_{\min}. \tag{12}$$

In most applications, the signal to be modeled contains white Gaussian noise. White noise tends to flatten the signal and distort the correlation function. If x_k is contaminated by white Gaussian noise v_k with zero mean and variance σ_v^2, the input to the prediction error filter, which is denoted as z_k, becomes

$$z_k = x_k + v_k. \tag{13}$$

Let $Z_k = [z_{k-1} \quad z_{k-2} \quad \cdots \quad z_{k-p}]^T$ be the noisy input vector. Equation (3) now becomes

$$W_{k+1} = W_k + \mu Z_k e_k^n \tag{14}$$

where

$$e_k^n = z_k - Z_k^T W_k \tag{15}$$

is the prediction error based on noisy inputs. Define $R_z = E\{Z_k Z_k^T\} = R + \sigma_v^2 I$ and $P_z = E\{z_k Z_k\} = P$. From (14), we find that the mean weight vector converges to

$$W_z = R_z^{-1} P_z = (R + \sigma_v^2 I)^{-1} P. \tag{16}$$

Comparing (16) with (2), it is clear that the optimal solution is biased. This bias is due to the extra terms in the diagonal of the correlation matrix of Z_k.

III. The γ-LMS Filter

A. Formulation and Mean Convergence

To solve the bias problem, Treichler suggested the γ-LMS algorithm [9], which is described as follows:

$$W_{k+1} = \gamma W_k + \mu Z_k e_k^n. \tag{17}$$

If σ_v^2 is known and

$$\gamma = 1 + \mu\sigma_v^2 \tag{18}$$

$E\{W_k\}$ will converge to W^*. This can easily be shown. Taking the expectation from (17), we have

$$\begin{aligned} E\{W_{k+1}\} = E\{W_k\} + \mu E\{(\sigma_v^2 I - Z_k Z_k^T) W_k\} \\ + \mu E\{z_k Z_k\}. \end{aligned} \tag{19}$$

Let W_k and Z_k be independent. Then, $E\{Z_k Z_k^T W_k\} = E\{Z_k Z_k^T\} E\{W_k\}$. In the steady state, (19) becomes

$$E\{W_k\} = (R_z - \sigma_v^2 I)^{-1} P_z = R^{-1} P \tag{20}$$

which is the Wiener solution. Apparently, the γ-LMS algorithm utilizes the noise variance σ_v^2 to cancel the extra terms in the diagonal of the correlation matrix of Z_k.

B. The Second-Order Statistics

In [9], Treichler did not provide the second-order statistics for γ-LMS. Here, we derive them in detail. First, (17) can be rewritten in terms of weight-error vectors

$$\begin{aligned}\boldsymbol{\epsilon}_{k+1} &= \big[I+\mu\big(\sigma_v^2 I - Z_k Z_k^T\big)\big]\boldsymbol{\epsilon}_k \\ &\quad + \mu\big[Z_k z_k + \big(\sigma_v^2 I - Z_k Z_k^T\big)W^*\big]. \end{aligned} \tag{21}$$

Thus, the correlation matrix of the weight-error vector K_{k+1} can be evaluated by

$$\begin{aligned} K_{k+1} &= E\big\{\boldsymbol{\epsilon}_{k+1}\boldsymbol{\epsilon}_{k+1}^T\big\} \\ &= E\Big\{\big[I+\mu\big(\sigma_v^2 I - Z_k Z_k^T\big)\big]\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T \\ &\qquad \times \big[\big(I+\mu\big(\sigma_v^2 I - Z_k Z_k^T\big)\big]\Big\} \\ &\quad + \mu^2 E\Big\{\big[Z_k z_k + \big(\sigma_v^2 I - Z_k Z_k^T\big)W^*\big] \\ &\qquad \times \big[z_k Z_k^T + W^{*T}\big(\sigma_v^2 I - Z_k Z_k^T\big)\big]\Big\}. \end{aligned} \tag{22}$$

Substituting (2) into (22), we have

$$\begin{aligned} K_{k+1} &= \big(I + 2\mu\sigma_v^2 + \mu^2\sigma_v^4\big)K_k \\ &\quad - \big(\mu + \mu^2\sigma_v^2\big)\big(K_k R_z + R_z K_k\big) - \sigma_v^4 W^* W^{*T} \\ &\quad + \mu^2 E\big\{Z_k Z_k^T \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T Z_k Z_k^T\big\} + \mu^2 E\big\{Z_k z_k z_k Z_k^T\big\} \\ &\quad - \mu^2 E\big\{Z_k z_k W^{*T} Z_k Z_k^T\big\} \\ &\quad - \mu^2 E\big\{Z_k Z_k^T W^* z_k Z_k^T\big\} \\ &\quad + \mu^2 E\big\{Z_k Z_k^T W^* W^{*T} Z_k Z_k^T\big\}. \end{aligned} \tag{23}$$

Equation (23) involves fourth-order moments of the input signal. These high-order moments can be evaluated by using the Gaussian moment factoring theorem. Let a_1, a_2, a_3, and a_4 denote four samples of a real Gaussian process with zero mean. The Gaussian moment factoring theorem states that

$$\begin{aligned} E\{a_1 a_2 a_3 a_4\} &= E\{a_1 a_2\}E\{a_3 a_4\} + E\{a_1 a_3\}E\{a_2 a_4\} \\ &\quad + E\{a_1 a_4\}E\{a_2 a_3\}. \end{aligned} \tag{24}$$

Denote the five expectation terms in (23) as A_k, B_k, C_k, D_k, and F_k. Assuming x_k is a Gaussian process, invoking the fundamental assumption [6], and using (24), we have

$$\begin{aligned} A_k &= \{A_{k,ij}\} = E\big\{Z_k Z_k^T \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T Z_k Z_k^T\big\} \\ &= \left\{\sum_{l=1}^{p}\sum_{m=1}^{p} E\{z_{k-i} z_{k-l}\epsilon_{k,l}\epsilon_{k,m} z_{k-m} z_{k-j}\}\right\} \\ &= \left\{\sum_{l=1}^{p}\sum_{m=1}^{p} E\{z_{k-i} z_{k-l} z_{k-m} z_{k-j}\}E\{\epsilon_{k,l}\epsilon_{k,m}\}\right\} \\ &= \left\{\sum_{l=1}^{p}\sum_{m=1}^{p} E\{z_{k-i} z_{k-l}\}E\{z_{k-m} z_{k-j}\}E\{\epsilon_{k,l}\epsilon_{k,m}\}\right\} \\ &\quad + \left\{\sum_{l=1}^{p}\sum_{m=1}^{p} E\{z_{k-i} z_{k-m}\}E\{z_{k-l} z_{k-j}\}E\{\epsilon_{k,l}\epsilon_{k,m}\}\right\} \\ &\quad + \left\{E\{z_{k-i} z_{k-j}\}\sum_{l=1}^{p}\sum_{m=1}^{p} E\{z_{k-l} z_{k-m}\}E\{\epsilon_{k,l}\epsilon_{k,m}\}\right\} \\ &= 2R_z K_k R_z + R_z \mathrm{tr}(R_z K_k) \end{aligned} \tag{25}$$

$$\begin{aligned} B_k &= \{B_{k,ij}\} = E\big\{Z_k z_k z_k Z_k^T\big\} \\ &= \{E\{z_{k-i} z_k z_k z_{k-j}\}\} \\ &= \{2E\{z_{k-i} z_k\}E\{z_k z_{k-j}\}\} + \{E\{z_{k-i} z_{k-j}\}E\{z_k z_k\}\} \\ &= 2P_z P_z^T + R_z\sigma_z^2 = 2PP^T + R_z\big(\sigma_x^2 + \sigma_v^2\big) \end{aligned} \tag{26}$$

and

$$\begin{aligned} C_k &= \{C_{k,ij}\} = E\big\{Z_k z_k W^{*T} Z_k Z_k^T\big\} \\ &= \left\{\sum_{l=1}^{p} E\{z_{k-i} z_k w_l^* z_{k-l} z_{k-j}\}\right\} \\ &= \left\{E\{z_{k-i} z_k\}\sum_{l=1}^{p} w_l^* E\{z_{k-l} z_{k-j}\}\right\} \\ &\quad + \left\{E\{z_k z_{k-j}\}\sum_{l=1}^{p} w_l^* E\{z_{k-i} z_{k-l}\}\right\} \\ &\quad + \left\{E\{z_{k-i} z_{k-j}\}\sum_{l=1}^{p} w_l^* E\{z_k z_{k-l}\}\right\} \\ &= P_z W^{*T} R_z + R_z W^* P_z^T + R_z W^{*T} P_z \\ &= PW^{*T} R + RW^* P^T + \sigma_v^2(PW^{*T} + W^* P^T) \\ &\quad + R_z W^{*T} P. \end{aligned} \tag{27}$$

Substituting (2) into (27), we have

$$C_k = 2PP^T + \sigma_v^2(PW^{*T} + W^* P^T) + R_z W^{*T} P. \tag{28}$$

Note that D_k is equal to the transpose of C_k

$$\begin{aligned} D_k &= \{D_{k,ij}\} = E\big\{Z_k Z_k^T W^* z_k Z_k^T\big\} = C_k^T \\ &= 2PP^T + \sigma_v^2(PW^{*T} + W^* P^T) + R_z W^{*T} P. \end{aligned} \tag{29}$$

F_k can be obtained by a procedure similar to (25).

$$\begin{aligned} F_k &= \{F_{k,ij}\} = E\big\{Z_k Z_k^T W^* W^{*T} Z_k Z_k^T\big\} \\ &= 2R_z W^* W^{*T} R_z + R_z \mathrm{tr}\big(R_z W^* W^{*T}\big) \\ &= 2\big(R + \sigma_v^2 I\big)W^* W^{*T}\big(R + \sigma_v^2 I\big) \\ &\quad + R_z \mathrm{tr}\big[\big(R + \sigma_v^2 I\big)W^* W^{*T}\big]. \end{aligned} \tag{30}$$

Substituting (2) into (30) and noting that $\mathrm{tr}(W^* W^{*T}) = W^{*T} W^*$, we can rewrite F_k as follows:

$$\begin{aligned} F_k &= 2PP^T + 2\sigma_v^2\big(PW^{*T} + W^* P^T + \sigma_v^2 W^* W^{*T}\big) \\ &\quad + R_z\big(W^{*T} P + \sigma_v^2 W^{*T} W^*\big). \end{aligned} \tag{31}$$

Finally, substituting (25), (26), (28), (29), and (31) into (23), we obtain the time evolution of the correlation matrix of the weight-error vector

$$\begin{aligned} K_{k+1} &= \big(1 + 2\mu\sigma_v^2 + \mu^2\sigma_v^4\big)K_k \\ &\quad - \big(\mu + \mu^2\sigma_v^2\big)\big(K_k R_z + R_z K_k\big) + 2\mu^2 R_z K_k R_z \\ &\quad + \mu^2 R_z \mathrm{tr}(R_z K_k) + \mu^2 R_z J_{\min}^{\gamma} + \mu^2\sigma_v^4 W^* W^{*T} \end{aligned} \tag{32}$$

where $J_{\min}^{\gamma}$ is the MSE yielded by using W^* on noisy inputs and is defined as

$$\begin{aligned} J_{\min}^{\gamma} &= E\big\{\big(z_k - Z_k^T W^*\big)^2\big\} \\ &= \sigma_x^2 - W^{*T} P + \sigma_v^2 W^{*T} W^* + \sigma_v^2 \\ &= J_{\min} + \sigma_v^2 W^{*T} W^* + \sigma_v^2. \end{aligned} \tag{33}$$

To simplify (32), we can rotate coordinates of X_k. The particular coordinate rotation is described by

$$Q^T R Q = \Lambda \tag{34}$$

where $\Lambda = \mathrm{diag}([\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p])$ is a diagonal matrix consisting of the eigenvalues of the correlation matrix R, and Q is the unitary matrix consisting of the eigenvectors associated with these eigenvalues. Note that $Q^T Q = I$. Furthermore, let

$$Q^T K_k Q = S_k. \tag{35}$$

Using the transformation described by (34) and (35), we can rewrite the recursive equation (32) as follows:

$$\begin{aligned} S_{k+1} = {} & \big(1 + 2\mu\sigma_v^2 + \mu^2\sigma_v^4\big) S_k \\ & - 2\big(\mu + \mu^2\sigma_v^2\big)\big(\Lambda + \sigma_v^2 I\big) S_k + 2\mu^2\big(\Lambda + \sigma_v^2 I\big)^2 S_k \\ & + \mu^2\big(\Lambda + \sigma_v^2 I\big) J_k^{\gamma,\mathrm{ex}} + \mu^2\big(\Lambda + \sigma_v^2 I\big) J_{\min}^{\gamma} \\ & + \mu^2\sigma_v^4 Q^T W^* W^{*T} Q \end{aligned} \tag{36}$$

where $J_k^{\gamma,\mathrm{ex}}$ is defined as

$$J_k^{\gamma,\mathrm{ex}} = \mathrm{tr}(R_z K_k) = \mathrm{tr}\big[\big(\Lambda + \sigma_v^2 I\big) S_k\big]. \tag{37}$$

Let $h_i, i = 1, 2, \ldots, p$ denote the diagonal term of $Q^T W^* W^{*T} Q$ and $s_{k,i}$ the diagonal term of S_k. Then

$$\begin{aligned} s_{k+1,i} = {} & s_{k,i} - 2\mu\lambda_i s_{k,i} + \mu^2\big[2\big(\lambda_i + \sigma_v^2\big)\lambda_i s_{k,i} + \sigma_v^4 s_{k,i} \\ & + \big(\lambda_i + \sigma_v^2\big) J_k^{\gamma,\mathrm{ex}} + \big(\lambda_i + \sigma_v^2\big) J_{\min}^{\gamma}\big] + \mu^2\sigma_v^4 h_i. \end{aligned} \tag{38}$$

In the steady state, (38) can be further simplified. If the step size is small, $J_k^{\gamma,\mathrm{ex}}$ is much smaller than $J_{\min}^{\gamma}$. Thus, $(\lambda_i + \sigma_v^2) J_k^{\gamma,\mathrm{ex}}$ can be ignored in comparison with $(\lambda_i + \sigma_v^2) J_{\min}^{\gamma}$. From (37), we note that $J_k^{\gamma,\mathrm{ex}} = \sum_{j=1}^{p} (\lambda_j + \sigma_v^2) s_{k,j}$. Therefore, $(\lambda_i + \sigma_v^2) J_k^{\gamma,\mathrm{ex}}$ is larger than $\sigma_v^4 s_{k,i}$. This implies that $\sigma_v^4 s_{k,i}$ can be ignored. Moreover

$$2\big(\lambda_i + \sigma_v^2\big)\lambda_i s_{k,i} \le 2\big(\lambda_i + \sigma_v^2\big) J_k^{\gamma,\mathrm{ex}}. \tag{39}$$

The equality holds only when the signal is a first-order AR process. Thus, $2(\lambda_i + \sigma_v^2)\lambda_i s_{k,i}$ can also be ignored. Finally, we have the steady state $s_{k,i}$ as follows:

$$s_{\infty,i} \approx \frac{1}{2}\mu\left(1 + \frac{\sigma_v^2}{\lambda_i}\right) J_{\min}^{\gamma} + \frac{1}{2}\mu\frac{\sigma_v^4 h_i}{\lambda_i}. \tag{40}$$

Therefore, the steady-state MSD becomes

$$\begin{aligned} M_\infty = \mathrm{tr}(K_\infty) = \mathrm{tr}(S_\infty) &= \sum_{i=1}^{p} s_{\infty,i} \\ &\approx \frac{1}{2}\mu p J_{\min}^{\gamma} + \frac{1}{2}\mu\sigma_v^2 \mathrm{tr}(R^{-1}) J_{\min}^{\gamma} \\ &\quad + \frac{1}{2}\mu\sigma_v^4 \mathrm{tr}(R^{-1} W^* W^{*T}). \end{aligned} \tag{41}$$

Equation (41) shows that the MSD is affected by the noise variance and optimal weights. When the noise power increases, the MSD of the γ-LMS algorithm will also increase.

IV. The ρ-LMS Filter

A. Formulation

From (14) and (15), we know that the weight vector of the LMS filter is adapted by the noisy input vector Z_k and the prediction error based on Z_k, e_k^n. This causes the weight vector to converge to a biased solution. To reduce the effect of noise, we can first estimate the noise-free input x_k and noise-free prediction error e_k and then use the estimates in the LMS algorithm. Since the estimates of x_k and e_k will contain less noise, the LMS algorithm will give better performance.

Given an observation sequence $Z^k = \{z_1 \; z_2 \; \cdots \; z_k\}$, the optimal estimate of x_k based on the observation up to k is the conditional mean of x_k, which is

$$\hat{x}_k = E\big\{x_k \,|\, Z^k\big\}. \tag{42}$$

From Bayes' law and (13), the *a posterior* density function $p(x_k \,|\, Z^k)$ can be expanded as follows:

$$\begin{aligned} p\big(x_k \,|\, Z^k\big) &= p\big(x_k \,|\, z_k, Z^{k-1}\big) = \frac{p\big(x_k, z_k \,|\, Z^{k-1}\big)}{p\big(z_k \,|\, Z^{k-1}\big)} \\ &= \frac{p\big(x_k \,|\, Z^{k-1}\big) p\big(z_k \,|\, x_k\big)}{p\big(z_k \,|\, Z^{k-1}\big).} \end{aligned} \tag{43}$$

The density $p(z_k \,|\, x_k)$ can be determined by $p(v_k)$ and

$$\begin{aligned} p\big(z_k \,|\, Z^{k-1}\big) &= \int p\big(z_k \,|\, x_k, Z^{k-1}\big) p\big(x_k \,|\, Z^{k-1}\big) dx_k \\ &= \int p\big(x_k \,|\, Z^{k-1}\big) p\big(z_k \,|\, x_k\big) dx_k. \end{aligned} \tag{44}$$

Note that without making any assumptions, the recursive estimate of x_k is almost impossible. As in [14], we assume that $p(x_k \,|\, Z^{k-1})$ is Gaussian. Define

$$\mu_k = E\big\{x_k \,|\, Z^{k-1}\big\} \tag{45}$$

and

$$n_k = x_k - \mu_k, \quad \sigma_{n_k}^2 = E\big\{n_k^2\big\}. \tag{46}$$

It has been shown [14] that the conditional mean of x_k can be written as

$$\hat{x}_k = \mu_k + \sigma_{n_k}^2 g(z_k) \tag{47}$$

where $g(z_k)$ is the score function of $p(z_k \,|\, Z^{k-1})$, i.e.,

$$g(z_k) = -\frac{\partial p\big(z_k \,|\, Z^{k-1}\big)}{\partial z_k}\big[p\big(z_k \,|\, Z^{k-1}\big)\big]^{-1}. \tag{48}$$

However, without the signal model, we cannot find μ_k and $\sigma_{n_k}^2$. For the time being, we assume that $W_k \approx W^*$ and $E\{x_{k-i} \,|\, Z^{k-1}\} \approx E\{x_{k-i} \,|\, Z^{k-i}\}$ for $i = 1, \ldots, p$. Let $\hat{X}_k = [\hat{x}_{k-1} \; \hat{x}_{k-2} \; \cdots \; \hat{x}_{k-p}]$. From (45), we have

$$\begin{aligned} \mu_k &= E\big\{X_k^T W^* + d_k \,|\, Z^{k-1}\big\} \\ &= E\big\{\big[x_{k-1} \,|\, Z^{k-1}, x_{k-2} \,|\, Z^{k-1}, \ldots, x_{k-p} \,|\, Z^{k-1}\big]\big\} W^* \\ &\approx E\big\{\big[x_{k-1} \,|\, Z^{k-1}, x_{k-2} \,|\, Z^{k-2}, \ldots, x_{k-p} \,|\, Z^{k-p}\big]\big\} W_k \\ &= \hat{X}_k^T W_k \triangleq \hat{\hat{x}}_k. \end{aligned} \tag{49}$$

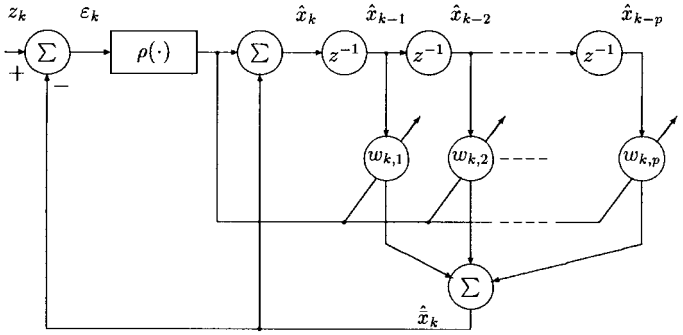

Fig. 2. ρ-LMS prediction error filter.

In (49), we have used the property that d_k is white noise and independent of Z^{k-1}. Thus, $\hat{\bar{x}}_k$ is the prediction based on $\hat{X}_k$ and W_k. We can also approximate $\sigma_{n_k}^2$ using the the conditional prediction error variance $\sigma_{\xi_k}^2$ in which

$$\xi_k = x_k - \hat{\bar{x}}_k, \quad \sigma_{\xi_k}^2 = E\{\xi_k^2\}. \tag{50}$$

Thus, (47) can be written as follows:

$$\hat{x}_k = \hat{\bar{x}}_k + \sigma_{\xi_k}^2 g(z_k). \tag{51}$$

Next, we consider the optimal estimate of e_k.

$$\begin{aligned}\hat{e}_k &= E\{e_k \mid Z^k\} = E\{x_k \mid Z^k\} - E\{\bar{x}_k \mid Z^k\} \\ &= \hat{x}_k - E\{\bar{x}_k \mid Z^k\}.\end{aligned} \tag{52}$$

Since it is difficult to obtain $E\{\bar{x}_k \mid Z^k\}$, we use $\hat{\bar{x}}_k$ to replace it. From (51) and (52), we have

$$\hat{e}_k \approx \hat{x}_k - \hat{\bar{x}}_k = \sigma_{\xi_k}^2 g(z_k). \tag{53}$$

Define the noisy prediction error ε_k as

$$\varepsilon_k = z_k - \hat{\bar{x}}_k = \xi_k + v_k. \tag{54}$$

Equation (53) can be rewritten as a function of ε_k

$$\hat{e}_k \approx \sigma_{\xi_k}^2 g(\varepsilon_k + \hat{\bar{x}}_k) = \rho(\varepsilon_k). \tag{55}$$

The function $\rho(\cdot)$ can be seen as the filtering operation on noisy prediction error ε_k. Substituting (55) into (51), we have

$$\hat{x}_k = \hat{\bar{x}}_k + \rho(\varepsilon_k). \tag{56}$$

Utilizing $\hat{x}_k$ as the input to the LMS prediction error filter and $\rho(\varepsilon_k)$ as the error signal to adjust filter weights, we obtain the weight-update equation for the ρ-LMS filter as follows:

$$W_{k+1} = W_k + \mu \hat{X}_k \rho(\varepsilon_k). \tag{57}$$

For comparison, we list the weight-update equation for the conventional LMS filter below.

$$W_{k+1} = W_k + \mu Z_k e_k^n \tag{58}$$

The ρ-LMS algorithm replaces the noisy input z_k with $\hat{x}_k$, which is an estimate of the noise-free input x_k and the prediction error based on noisy inputs e_k^n with $\rho(\varepsilon_k)$, which is an estimate of the noise-free prediction error e_k. The structure of the ρ-LMS filter is illustrated in Fig. 2.

B. The ρ-LMS Filter for White Gaussian Noise

The density function of Gaussian noise v_k with variance σ_v^2 is given by

$$p(v_k) = N_{\sigma_v^2}(v_k) = \left(2\pi\sigma_v^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{v_k^2}{2\sigma_v^2}\right\}. \tag{59}$$

Thus, from (44), we can express $p(z_k \mid Z^{k-1})$ as

$$\begin{aligned}p(z_k \mid Z^{k-1}) &= N_{\sigma_{\xi_k}^2 + \sigma_v^2}(z_k - \hat{\bar{x}}_k) \\ &= \left[2\pi\left(\sigma_{\xi_k}^2 + \sigma_v^2\right)\right]^{-\frac{1}{2}} \exp\left\{-\frac{(z_k - \hat{\bar{x}}_k)^2}{2\left(\sigma_{\xi_k}^2 + \sigma_v^2\right)}\right\}.\end{aligned} \tag{60}$$

From (48), the score function of (60) is found to be

$$g(z_k) = \frac{z_k - \hat{\bar{x}}_k}{\sigma_{\xi_k}^2 + \sigma_v^2}. \tag{61}$$

Thus, the estimate of the noise-free prediction error in (55) is

$$\rho(\varepsilon_k) = \varepsilon_k \frac{\sigma_{\xi_k}^2}{\sigma_{\xi_k}^2 + \sigma_v^2}. \tag{62}$$

It is not surprising that the filter function is linear since the noise is Gaussian. Define the filter gain as

$$g_k = \frac{\sigma_{\xi_k}^2}{\sigma_{\xi_k}^2 + \sigma_v^2}. \tag{63}$$

We can rewrite (56) and (57) as

$$\hat{x}_k = \hat{\bar{x}}_k + g_k \varepsilon_k \tag{64}$$

$$W_{k+1} = W_k + \mu g_k \hat{X}_k \varepsilon_k. \tag{65}$$

C. The Convergence Analysis

From (63) and (64), we find that $\hat{X}_k$ is nonstationary; it is, therefore, difficult to analyze the transient behavior of the ρ-LMS algorithm. In what follows, we will concentrate on analyzing the steady-state behavior of the ρ-LMS algorithm. In the steady state, we assume that ξ_k approaches a stationary white process with a constant variance σ_ξ^2. This implies that g_k, which is denoted as g, is constant. From (4), (54), and (64), $\hat{X}_k$ and ε_k can be written as follows:

$$\hat{X}_k = X_k - (1-g)\Xi_k + gV_k \tag{66}$$

$$\varepsilon_k = e_k + (1-g)\Xi_k^T W_k - gV_k^T W_k + v_k \tag{67}$$

where $\Xi_k = [\xi_{k-1} \ \xi_{k-2} \ \cdots \ \xi_{k-p}]^T$ and $V_k = [v_{k-1} \ v_{k-2} \ \cdots \ v_{k-p}]^T$. Thus, the term of $\hat{X}_k \varepsilon_k$ can be expanded to

$$\begin{aligned}\hat{X}_k \varepsilon_k = {} & X_k e_k - (1-g)\Xi_k e_k + gV_k e_k \\ & + (1-g)X_k \Xi_k^T W_k - (1-g)^2 \Xi_k \Xi_k^T W_k \\ & + g(1-g)V_k \Xi_k^T W_k - gX_k V_k^T W_k \\ & + g(1-g)\Xi_k V_k^T W_k - g^2 V_k V_k^T W_k \\ & + X_k v_k - (1-g)\Xi_k v_k + gV_k v_k.\end{aligned} \tag{68}$$

To make the analysis mathematically tractable, we make the following assumptions:

1) W_k is independent of X_k, Ξ_k, and V_k.
2) Ξ_k is independent of e_k and V_k.
3) ξ_k is independent of $\hat{\bar{x}}_k$.

Note that this is an extension of the fundamental assumption in the analysis of the conventional LMS algorithm [6]. Applying Assumptions 1 and 2, we can write the expectation of (68) as

$$E\{\hat{X}_k\varepsilon_k\} = E\{X_k e_k\} + (1-g)E\{X_k\Xi_k^T\}E\{W_k\} - (1-g)^2\sigma_\xi^2 E\{W_k\} - g^2\sigma_v^2 E\{W_k\}. \quad (69)$$

$E\{X_k\Xi_k^T\}$ can be evaluated using the relation $x_k = \hat{\bar{x}}_k + \xi_k$ and Assumption 3. Then, (69) is reduced to

$$\begin{aligned} E\{\hat{X}_k\varepsilon_k\} &= E\{X_k e_k\} + \big[(1-g)\sigma_\xi^2 \\ &\quad - (1-g)^2\sigma_\xi^2 - g^2\sigma_v^2\big]E\{W_k\} \\ &= E\{X_k e_k\} + g\big[(1-g)\sigma_\xi^2 - g\sigma_v^2\big]E\{W_k\}. \end{aligned} \quad (70)$$

Using the relation $g = \sigma_\xi^2/(\sigma_\xi^2+\sigma_v^2)$, we find $(1-g)\sigma_\xi^2 - g\sigma_v^2 = 0$. Thus, we have

$$E\{\hat{X}_k\varepsilon_k\} = E\{X_k e_k\}. \quad (71)$$

The expectation of (65) now becomes

$$\begin{aligned} E\{W_{k+1}\} &= E\{W_k\} + \mu g E\{\hat{X}_k\varepsilon_k\} \\ &= E\{W_k\} + \mu g E\{X_k e_k\}. \end{aligned} \quad (72)$$

Note that (72) is simply the recursive equation for the mean weight vector of the conventional LMS algorithm (with noise-free input x_k and step size μg). Thus, we can say that the mean weight vector of the ρ-LMS algorithm converges to the Wiener solution $R^{-1}P$. The stability condition for the step size is then

$$0 < \mu < \frac{2}{g\lambda_{\max}} \quad (73)$$

where $\lambda_{\max}$ is the maximum eigenvalue of R.

In the rest of this subsection, we will consider the MSD of the ρ-LMS filter. In the steady state, the filter gain approaches a constant and can be absorbed into the step size. Equation (65) then becomes

$$W_{k+1} = W_k + \mu\hat{X}_k\varepsilon_k. \quad (74)$$

Rewriting (74) in terms of the weight-error vector, we have

$$\boldsymbol{\epsilon}_{k+1} = \big[I - \mu\hat{X}_k\hat{X}_k^T\big]\boldsymbol{\epsilon}_k + \mu\hat{X}_k\varepsilon_{o,k} \quad (75)$$

where

$$\varepsilon_{o,k} = z_k - \hat{X}_k^T W^*. \quad (76)$$

Assume that $\hat{x}_k$ is a stationary Gaussian process and $\hat{X}_k$ and $\varepsilon_{o,k}$ are uncorrelated. From (75), the correlation matrix of the weight-error vector is derived as

$$\begin{aligned} K_{k+1} &= E\big\{\big(I - \mu\hat{X}_k\hat{X}_k^T\big)\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T\big(I - \mu\hat{X}_k\hat{X}_k^T\big)\big\} \\ &\quad + \mu^2 E\big\{\hat{X}_k\varepsilon_o\varepsilon_{o,k}\hat{X}_k^T\big\} \\ &= K_k - \mu(K_k R_{\hat{X}} + R_{\hat{X}}K_k) \\ &\quad + \mu^2 E\big\{\hat{X}_k\hat{X}_k^T\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T\hat{X}_k\hat{X}_k^T\big\} + \mu^2 R_{\hat{X}}E\{\varepsilon_{o,k}^2\} \end{aligned} \quad (77)$$

where $R_{\hat{X}} = E\{\hat{X}_k\hat{X}_k^T\}$. The third term on the right-hand side of (77) can be expanded by the Gaussian moment factoring theorem. From (25), we have

$$E\big\{\hat{X}_k\hat{X}_k^T\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^T\hat{X}_k\hat{X}_k^T\big\} = 2R_{\hat{X}}K_k R_{\hat{X}} + R_{\hat{X}}\mathrm{tr}(R_{\hat{X}}K_k). \quad (78)$$

Substitute (78) into (77). We obtain

$$\begin{aligned} K_{k+1} &= K_k - \mu(K_k R_{\hat{X}} + R_{\hat{X}}K_k) + 2\mu^2 R_{\hat{X}}K_k R_{\hat{X}} \\ &\quad + \mu^2 R_{\hat{X}}\mathrm{tr}(R_{\hat{X}}K_k) + \mu^2 R_{\hat{X}}J_{\min}^f \end{aligned} \quad (79)$$

where $J_{\min}^f = E\{\varepsilon_{o,k}^2\}$. We can find $J_{\min}^f$ by using (76) and (67).

$$\varepsilon_{o,k} = e_{o,k} + (1-g)\Xi_k^T W^* - gV_k^T W^* + v_k \quad (80)$$

where $e_{o,k} = x_k - X_k^T W^*$. Squaring both sides of (80) and taking expectation of the result, we have

$$J_{\min}^f = J_{\min} + (1-g)^2\sigma_\xi^2 W^{*T}W^* + g^2\sigma_v^2 W^{*T}W^* + \sigma_v^2. \quad (81)$$

Using the relation $(1-g)\sigma_\xi^2 = g\sigma_v^2$, we can rewrite (81) as

$$J_{\min}^f = J_{\min} + g\sigma_v^2 W^{*T}W^* + \sigma_v^2. \quad (82)$$

Similar to the derivation of the MSD of the γ-LMS algorithm in (32), we can rotate the coordinates of $\hat{X}_k$ such that $R_{\hat{X}}$ becomes diagonal. Let $s_{k,i}$ be the diagonal term of K_k after rotation. From (79), the equation for updating $s_{k,i}$ can be written as

$$\begin{aligned} s_{k+1,i} &= s_{k,i} - 2\mu\alpha_i s_{k,i} + 2\mu^2\alpha_i^2 s_{k,i} \\ &\quad + \mu^2\alpha_i J_k^{f,\mathrm{ex}} + \mu^2\alpha_i J_{\min}^f \end{aligned} \quad (83)$$

where α_i's are eigenvalues of $R_{\hat{X}}$, and $J_k^{f,\mathrm{ex}} = \mathrm{tr}(R_{\hat{X}}K_k)$. The third and fourth terms on the right-hand side of (83) can be ignored if μ is small. Therefore, the MSD of the ρ-LMS filter in the steady state can be approximated by

$$M_\infty \approx \frac{1}{2}\mu p J_{\min}^f. \quad (84)$$

Comparing (82) with (33) and noting that $g < 1$, we find that $J_{\min}^f$ is smaller than $J_{\min}^\gamma$. Thus, (84) is smaller than the first term of (41). The second term of (41) is positive. Since R^{-1} is positive definite, $\mathrm{tr}(R^{-1}W^*W^{*T}) = \mathrm{tr}(W^{*T}R^{-1}W^*) > 0$. Thus, the third term in (41) is also positive. We conclude that for the same step size, the steady-state MSD of the ρ-LMS filter is smaller than that of the γ-LMS filter. In the next section, we will present experiments to show the accuracy of our theoretical results.

D. Practical Implementations

To use (65), the variance $\sigma_{\xi_k}^2$ in (63) must be estimated. From (54), we find that $\sigma_{\xi_k}^2 = \sigma_{\varepsilon_k}^2 - \sigma_v^2$. However, we cannot use this relation to estimate $\sigma_{\xi_k}^2$. The reason is explained as below. In (49) and (50), we use $\hat{\bar{x}}_k$ and $\sigma_{\xi_k}^2$ to

approximate μ and $\sigma_{n_k}^2$. Note that these approximations are based on two assumptions: $W_k \approx W^*$ and $E\{x_{k-i} \mid Z^{k-1}\} \approx E\{x_{k-i} \mid Z^{k-i}\}$ for $i = 1, \ldots, p$. In the transient state, these two assumptions are not valid. As a consequence, $\sigma_{\xi_k}^2$ is much larger than $\sigma_{n_k}^2$. Here, we develop another method to overcome this problem. Observe that the filter tries to make $\hat{\bar{x}}_k$ (i.e., ξ_k) as close to $\bar{x}_k$ (i.e., e_k) as possible. Thus, it will be reasonable to let

$$\sigma_{\xi_k}^2 \approx \sigma_{e_k}^2. \tag{85}$$

$\sigma_{e_k}^2$ can be derived using a procedure similar to $J_{\min}^f$. Assume that during the period $[k-p, k-1]$, g_k, and $\sigma_{\xi_k}^2$ change slowly and can be seen as constants. From (67), $\sigma_{\varepsilon_k}^2$ is then given by

$$\begin{aligned}\sigma_{\varepsilon_k}^2 = {} & E\{e_k^2\} + (1-g_{k-1})^2 \sigma_{\xi_{k-1}}^2 E\{W_k^T W_k\} \\ & + g_{k-1}^2 \sigma_v^2 E\{W_k^T W_k\} + \sigma_v^2\end{aligned} \tag{86}$$

where $\sigma_{\varepsilon_k}^2 = E\{\varepsilon_k^2\}$ and $\sigma_{\xi_{k-1}}^2 = E\{\xi_{k-1}^2\}$. Using the relations $(1-g_{k-1})\sigma_{\xi_{k-1}}^2 = g_{k-1}\sigma_v^2$, we can rewrite (86) as

$$\sigma_{\varepsilon_k}^2 = \sigma_{e_k}^2 + g_{k-1}\sigma_v^2 E\{W_k^T W_k\} + \sigma_v^2. \tag{87}$$

From (85), We obtain $\sigma_{\xi_k}^2$ as

$$\sigma_{\xi_k}^2 \approx \sigma_{\varepsilon_k}^2 - g_{k-1}\sigma_v^2 E\{W_k^T W_k\} - \sigma_v^2. \tag{88}$$

In practice, the expectation terms in (88) cannot be obtained. Thus, a fading-memory average is used to recursively estimate $\sigma_{\xi_k}^2$

$$\hat{\sigma}_{\xi_k}^2 = \beta \hat{\sigma}_{\xi_{k-1}}^2 + (1-\beta)(\varepsilon_k^2 - \hat{g}_{k-1}\sigma_v^2 W_k^T W_k - \sigma_v^2) \tag{89}$$

where $\hat{\sigma}_{\xi_k}^2$ is the estimate of $\sigma_{\xi_k}^2$, and $\hat{g}_{k-1} = \hat{\sigma}_{\xi_{k-1}}^2/(\hat{\sigma}_{\xi_{k-1}}^2 + \sigma_v^2)$ is the estimate of g_{k-1}. β is a forgetting factor and is chosen to be close to 1.

Finally, we summarize the whole algorithm for the linear ρ-LMS filter as follows:

Step 1) $\hat{\bar{x}}_k = \hat{X}_k^T W_k$.
Step 2) $\varepsilon_k = z_k - \hat{\bar{x}}_k$.
Step 3) $\hat{\sigma}_{\xi_k}^2 = \beta\hat{\sigma}_{\xi_{k-1}}^2 + (1-\beta)(\varepsilon_k^2 - \hat{g}_{k-1}\sigma_v^2 W_k^T W_k - \sigma_v^2)$.
Step 4) $\hat{g}_k = \hat{\sigma}_{\xi_k}^2/(\hat{\sigma}_{\xi_k}^2 + \sigma_v^2)$.
Step 5) $\hat{x}_k = \hat{\bar{x}}_k + \hat{g}_k\varepsilon_k$.
Step 6) $W_{k+1} = W_k + \mu\hat{g}_k\hat{X}_k\varepsilon_k$.
Step 7) $k = k+1$, and go to step 1.

V. Simulations

Computer simulations were carried out to evaluate the accuracy of our theoretical MSD's for the γ-LMS and ρ-LMS filters. In this study, two AR processes were used. One was a wideband signal obtained from

$$x_k = 0.6x_{k-1} - 0.25x_{k-2} + d_k \tag{90}$$

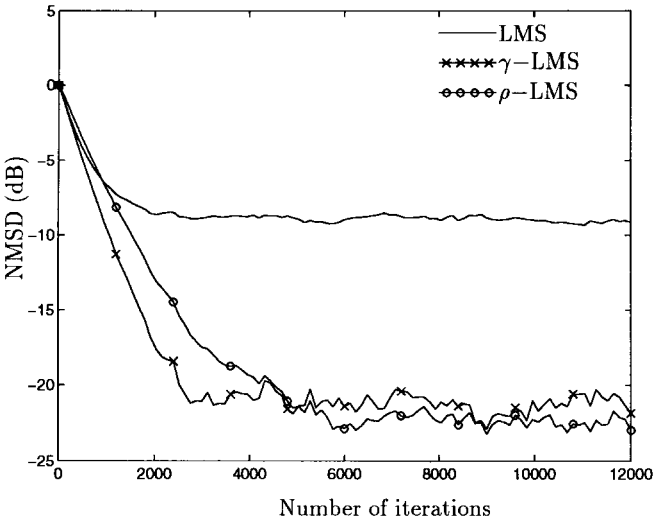

Fig. 3. Learning curve of NMSD for wide-band signal.

where d_k is white Gaussian noise. The corresponding poles are located at $0.3 \pm j0.4$. The other was a narrowband signal obtained from

$$x_k = 1.6x_{k-1} - 0.9536x_{k-2} + d_k \tag{91}$$

and having poles at $0.8 \pm j0.56$. In both cases, additive white Gaussian noise was used to contaminate x_k. The power of x_k was fixed at 10, and the input SNR was held at 5 dB. We defined the normalized-MSD as the performance criterion.

$$\text{NMSD} = 10\log_{10}\frac{M_k}{W^{*T}W^*} \text{ (dB)}. \tag{92}$$

To compare the NMSD at the same convergence speed, we used (74) instead of (65) to update the filter weights of the ρ-LMS algorithm. Fig. 3 shows the learning curves of the NMSD for the wideband signal from the LMS, γ-LMS, and ρ-LMS filters. The step size used here was 0.0002. Fig. 4 shows the results for the narrowband signal. The corresponding step size was 0.001. Both figures were obtained from an average of 50 runs with $\beta = 0.99$. In Fig. 3, we find that the ρ-LMS filter had only a slight performance improvement over the γ-LMS filter. This is because the signal was wideband making prediction difficult. For the narrowband signal, we see that the ρ-LMS filter had much better performance. In Fig. 4, we find that the NMSD of the ρ-LMS filter is about 8 dB lower than that of the γ-LMS filter. Note that the weight vector of the LMS filter converged to a biased solution; hence, it had poor performance in both cases.

As we know, the steady-state MSD is proportional to the step size, but the convergence rate is inversely proportional to the step size. To simultaneously include both into performance evaluation, we then define a comprehensive measure as follows:

$$\begin{aligned}\Omega = {} & \text{Steady state MSD} \\ & \times \text{Number of iterations to achieve convergence.}\end{aligned} \tag{93}$$

Thus, the smaller the Ω is, the better performance a filter has. We used the narrowband signal to perform simulations, and list

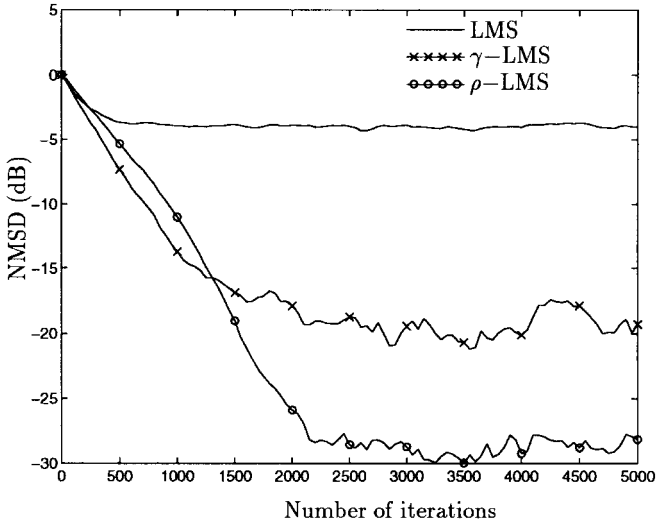

Fig. 4. Learning curve of NMSD for narrow-band signal.

TABLE I
Ω'S FOR THE γ-LMS and ρ-LMS FILTERS

Step size	$\gamma-$LMS	$\rho-$LMS
0.000125	83.95	13.80
0.00025	79.44	11.30
0.0005	77.84	9.784
0.001	81.64	9.884
0.002	119.8	18.74
0.004	125.8	33.78

TABLE II
EXPERIMENTAL AND THEORETICAL MSD's FOR THE γ-LMS FILTER

	Wide-band	Narrow-band
Experimental MSD	0.003469	0.007902
MSD in (41)	0.003370	0.007425
Ψ	2.94%	6.42%

Ω's for the γ-LMS and ρ-LMS filters in Table I. Each value was obtained from an average of 50 runs. We can see that Ω's for the ρ-LMS filter are significantly smaller than those of the γ-LMS filter in all cases. It is also worth mentioning that there is an optimal step size corresponding to the smallest Ω. For the simulations conducted here, the optimal step size is 0.0005.

The theoretical and the experimental MSD values of the γ-LMS and ρ-LMS filters are shown in Tables II and III, respectively. To evaluate the accuracy of the theoretical values, we define the error ratio Ψ, which is

$$\Psi = 100 \times \frac{\text{experimental value} - \text{theoretical value}}{\text{theoretical value}}\%. \tag{94}$$

For all cases, the step size used was 0.0002. The theoretical MSD values were computed by using (41) and (84) for the γ-LMS and the ρ-LMS filters, respectively. From these tables, we can see that the theoretical MSD values are close to the experimental ones. All Ψ's are below 10%.

VI. CONCLUSION

In AR modeling, if the input signal is corrupted by white Gaussian noise, the LMS prediction error filter will give biased coefficients. Treichler has suggested the γ-LMS filter to obtain unbiased solutions. In this paper, we applied the ρ-LMS filter proposed in [10] and [11] to the AR modeling problem. We first derived the second-order statistic of the γ-LMS filter, which is often used to measure the performance of adaptive filters. Then, using estimation theory, we derived the ρ-LMS filter and showed that the filter is linear when noise is Gaussian. We analyzed the first- and second-order statistics of the linear ρ-LMS filter and proved that it performs better than the γ-LMS filter. Experimental results demonstrate that our theoretical analysis is adequate.

Conventional approaches to the filtering problems involve two-stage operations. First, an algorithm is used to identify the signal model. Then, a filter is applied to perform the filtering operation. As a byproduct, the ρ-LMS filter can output filtered results for signals corrupted by white Gaussian noise. This is a significant advantage since identification and filtering are combined into a single filter. The ρ-LMS filter can be applied in many areas such as speech filtering, line enhancement, and active noise cancellation. Research in these directions is now underway.

TABLE III
EXPERIMENTAL AND THEORETICAL MSD's FOR THE ρ-LMS FILTER

	Wide-band	Narrow-band
Experimental MSD	0.002461	0.001593
MSD in (84)	0.002264	0.001504
Ψ	8.70%	5.92%

REFERENCES

[1] J. D. Gibson, B. Koo, and S. D. Grey, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1732–1741, Aug. 1991.
[2] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 197–210, June 1987.
[3] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[4] J. M. Pimbley, "Recursive autoregressive spectral estimation by minimization of free energy," *IEEE Trans. Signal Processing*, vol. 40, pp. 1518–1527, June 1992.
[5] A. V. Oppenheim, E. Weinstein, K. C. Zangi, M. Feder, and D. Gauger, "Single-sensor active noise cancellation," *IEEE Trans. Speech, Audio Processing*, vol. 2, pp. 285–290, Apr. 1994.
[6] S. S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[8] S. M. Kay, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 478–485, Oct. 1979.
[9] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53–62, Feb. 1979.
[10] R. Vijayan and H. V. Poor, "Nonlinear techniques for interference suppression in spread-spectrum system," *IEEE Trans. Commun.*, vol. 38, pp. 1060–1065, July 1990.
[11] L. A. Rusch and H. V. Poor, "Narrowband interference suppression in CDMA spread spectrum communications," *IEEE Trans. Commun.*, vol. 42, pp. 1969–1979, Feb./Mar./Apr. 1994.
[12] S. R. Kim and A. Efron, "Adaptive robust impulse noise filtering," *IEEE Trans. Signal Processing*, vol. 43, pp. 1855–1866, Aug. 1995.
[13] K. C. Ho and P. C. Ching, "Performance analysis of a split-path LMS adaptive filter for AR modeling," *IEEE Trans. Signal Processing*, vol. 40, pp. 1375–1382, June 1992.
[14] C. J. Masreliez, "Approximate non-Gaussian filtering with linear state and observation," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 107–110, Feb. 1975.

Wen-Rong Wu was born in Taiwan, R.O.C., in 1958. He received the B.S. degree in mechanical engineering from Tatung Institute of Technology, Taiwan, in 1980 and the M.S. degrees in mechanical and electrical engineering and the Ph.D. degree in electrical engineering from the State University of New York, Buffalo, in 1985, 1986, and 1989, respectively.

Since August 1898, he has been a faculty member with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include statistical signal processing and digital communications.

Po-Cheng Chen was born in Taiwan, R.O.C, in 1968. He received the B.S. and M.S. degrees in electrical engineering from Tatung Institute of Technology, Taiwan, in 1990 and 1992, respectively. Since 1992, he has been a Ph.D. student with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan.

His research interests include speech enhancement and adaptive filtering.