[3] D. Chazan, M. Zakai, and J. Ziv, "Improved lower bounds on signal parameter estimation," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 90–93, Jan. 1975.
[4] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. Commun.*, vol. COM-22, pp. 340–342, Mar. 1974.
[5] K. Bell, Y. Ephraim, and H. V. Trees, "Explicit Ziv-Zakai lower bounds for bearing estimation," *IEEE Trans. Signal Processing*, vol. 24, pp. 2810–2824, Nov. 1996.
[6] H. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

A Performance Bound for the LMS Estimator

Kevin J. Quirk, *Student Member, IEEE*,
Laurence B. Milstein, *Fellow, IEEE*, and
James R. Zeidler, *Fellow, IEEE*

***Abstract*—The least-mean-square (LMS) estimator is a nonlinear estimator with information dependencies spanning the entire set of data fed into it. The traditional analysis techniques used to model this estimator obscure these dependencies; to simplify the analysis they restrict the estimator to the finite set of data sufficient to span the length of its filter. Thus the finite Wiener filter is often considered a bound on the performance of the LMS estimator. Several papers have reported the performance of the LMS filter exceeding that of the finite Wiener filter. In this correspondence, we derive a bound on the LMS estimator which does not exclude the contributions from data outside its filter length. We give examples of this bound in cases where the LMS estimator outperforms the finite Wiener filter.**

***Index Terms*—Adaptive filter, LMS algorithm, multidimensional filtering, noise canceler, spectral factorization, Wiener filter.**

I. INTRODUCTION

The least-mean-square (LMS) adaptive filter was first introduced by Widrow. Since then, it has found widespread use in many applications [1], due in part to the simplicity of its implementation; it requires only a finite-impulse response (FIR) filter and a first-order weight update equation. This simplicity, however, belies what is actually a complex nonlinear estimator. A direct analysis of this estimator's performance does not appear to be feasible; therefore, attention has focused on restricting the statistics of the input processes to simplify this analysis. The "independence assumptions" approach, outlined in Section III of this correspondence, is the most common such method. Through invocation of this strict set of assumptions, the LMS estimator is modeled as the combination of a finite Wiener filter along with some "misadjustment noise" representing the difference between the converged LMS weights and the finite Wiener filter. This approach has been shown to be an effective way of analyzing the LMS estimator in many situations, even when the assumptions are not strictly met [2]. Based on the widespread success of this model, the finite Wiener filter is often mistakenly assumed to bound the performance of LMS estimators. Recently, several papers have reported cases where the performance of the LMS filter surpasses that of the finite Wiener filter. North *et al.* in [3] first reported a better bit-error rate performance from an LMS adaptive equalizer when used in the presence of a strong temporally correlated interferer than that which is achieved using the corresponding Wiener filter. Reuter and Zeidler in [4] followed this by analyzing the performance of the LMS equalizer in the presence of a sinusoidal interferer by using a transfer function approach to model the system as the combination of a steady-state filter and a time-varying filter. This model allowed for an analytical expression of the error which was used to solve for the optimal step size parameter. Douglas and Pan in [5] and Butterweck in [6] have also each presented methods for analyzing the LMS adaptive filter without use of the independence assumptions. The work in this correspondence extends the investigation of this phenomenon by providing an appropriate bound on the performance of the LMS estimator. We derive this bound in Section IV by bounding the performance of the LMS estimator by that of the optimal estimator for a class of signals without using the "independence assumptions." To solve for the optimal estimator we rely mostly on the wide-sense stationarity of the signals, applying spectral representations, and performing a spectral factorization. This bound, written for the LMS estimator in its most general form, is therefore valid for the four most common applications of the LMS estimator: equalization, noise cancellation, prediction, and system identification. We will then, in Section V, compare the performance of the LMS estimator with that of the optimal estimator in cases where the LMS estimator outperforms the finite Wiener filter.

II. BACKGROUND

Assume it is necessary to estimate the current value of a desired discrete time signal, $d[n]$. To form an estimate $\hat{d}[n]$, we have available to us the current and past values of a reference signal $u[n]$ which is correlated with the desired signal. We also have available to us the past error values of our estimate $e[n-1] = d[n-1] - \hat{d}[n-1]$. The LMS estimator uses these quantities in the filter structure given in Fig. 1 to produce an estimate of the desired signal $\hat{d}_{lms}[n]$.

The estimate is produced by passing the reference data through an L-tap FIR filter, where the filter weights are updated through the LMS weight update equation

$$\vec{\boldsymbol{w}}[n] = \vec{\boldsymbol{w}}[n-1] + \mu e^*[n-1]\vec{\boldsymbol{u}}_L[n-1]$$

and where

$$\vec{\boldsymbol{w}}[n] = [w_0[n], \cdots, w_{L-1}[n]]^T$$

and

$$\vec{\boldsymbol{u}}_L[n] = [u[n], \cdots, u[n-L+1]]^T$$

with the superscript T denoting the transpose operation.

The weight update equation is derived through a minimization of the mean-square error (MSE) between the desired signal and the LMS estimate, namely,

$$e_{lms}^2 = E\left\{\left|d[n] - \hat{d}_{\rm lms}[n]\right|^2\right\}. \quad (1)$$

Assuming that the initial weight vector at time $n = -\infty$ is the all-zero vector, we can write the LMS estimator as a nonlinear function of the

Manuscript received February 4, 1998; revised November 10, 1999. This work was supported in part by the NSF Industry/University Cooperative Research Center on Integrated Circuits and Systems at UCSD, and by the Focused Research Initiative on Wireless Multimedia Networks under Grant DAA HO4-95-1-0248. The material in this paper was presented in part at the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, May 12–15, 1998.

The authors are with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: kquirk@ucsd.edu; {milstein; zeidler}@ece.ucsd.edu).

Communicated by E. Soljanin, Associate Editor for Coding Techniques.
Publisher Item Identifier S 0018-9448(00)02901-1.

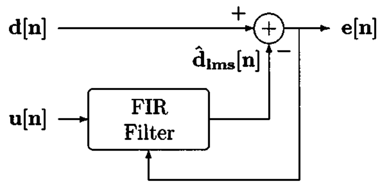

Fig. 1. LMS Estimator.

semi-infinite set of reference data, as well as of past values of the desired signal:

$$\hat{d}_{\text{lms}}[n] = \mu \sum_{i=-\infty}^{n-1} e[i]\vec{\boldsymbol{u}}_L^H[i]\vec{\boldsymbol{u}}_L[n] = f\left(u[n], \cdots, u[-\infty], d[n-1], \cdots, d[-\infty]\right), \quad (2)$$

where the superscript H represents the conjugate transpose.

A direct analysis of this estimator's performance does not appear to be feasible. The traditional approach to reducing the complexity of the analysis has been to restrict the statistics of input signals through a set of assumptions, collectively known as the "independence assumptions." The effect of these assumptions is to make the current filter-weight vector statistically independent of the current tap-data vector. This simplifies the MSE analysis and gives rise to a model of the LMS estimator weights as being those of the finite Wiener filter with an additional gradient noise. Instead of modeling the LMS estimator's performance, we will look at the bounding of its performance. First we will consider the performance bound produced through the strict "independence assumptions," and then look at the performance bound produced under a milder set of assumptions.

III. Independence Assumptions

The performance of any estimator can be bounded by that of the optimal estimator. The optimal MSE estimator is given by the mean of the desired signal, conditioned on the knowledge of all information available to the estimator [7, pp. 62–65]. Examining the equation for the LMS estimator (2), the optimal estimator is given by

$$\hat{d}_{\text{opt}}[n] = E\left\{d[n]|u[n], \cdots, u[-\infty], d[n-1], \cdots, d[-\infty]\right\}. \quad (3)$$

Actually solving for this estimator requires knowing the statistics of the signals $d[n]$ and $u[n]$. Under the "independence assumptions," the statistics of the signals are restricted through a set of assumptions. There are several derivations involving the convergence of the LMS estimator, and, depending on the measure of convergence, they differ as to the constraints which make up the independence assumptions. All of them assume the tap weight vector is independent of the tap data vector; however, depending on the derivation, they differ regarding the constraints on the desired signal [8], [9], [10, pp. 390–405]. Despite this, these analyses all result in the convergence of the LMS weights to those of the finite Wiener filter; we will therefore choose the constraints in [8] as the most straightforward to apply here:

1) the composite (desired signal, tap input vector) vectors $[d[n], \vec{\boldsymbol{u}}_L^T[n]]^T$, and

$$\left[d[n-1], \vec{\boldsymbol{u}}_L^T[n-1]\right]^T, \cdots, \left[d[-\infty], \vec{\boldsymbol{u}}_L^T[-\infty]\right]^T$$

are independent of each other;
2) $d[n]$ is dependent on $\vec{\boldsymbol{u}}_L[n]$;
3) $\vec{\boldsymbol{u}}_L[n]$ and $d[n]$ are mutually Gaussian.

Fig. 2. Independence assumptions bound.

These assumptions simplify the conditional mean, allowing for the optimal estimator to be obtained. Using Assumptions 1) and 2) reduces the conditional mean to

$$\hat{d}_{\text{ind}}[n] = E\left\{d[n]|u[n], \cdots, u[n-L+1]\right\}.$$

Condition 3) requires the optimal estimator to be linear, and is then given by

$$\hat{d}_{\text{ind}}[n] = \sum_{i=0}^{L-1} w_{\text{ind}}^*[i]u[n-i].$$

This is recognized as the finite Wiener filter operating on the L reference values $\{u[n], \cdots, u[n-L+1]\}$. Thus the performance of the LMS estimator, under these assumptions, can be bounded by that of the finite Wiener filter (Fig. 2), where the filter weights are given in terms of the autocorrelation matrix of the reference signal $\boldsymbol{R}$, and the crosscorrelation vector between the reference and desired signals $\vec{\boldsymbol{P}}$. Explicitly, the weights are

$$\boldsymbol{w}_{\text{ind}}[n] = \boldsymbol{R}^{-1}\vec{\boldsymbol{P}}$$

where

$$\boldsymbol{R} = E\left\{\vec{\boldsymbol{u}}_L[n]\vec{\boldsymbol{u}}_L^H[n]\right\}$$

and

$$\vec{\boldsymbol{P}} = E\left\{\vec{\boldsymbol{u}}_L[n]d^*[n]\right\}.$$

The MSE of the LMS estimator (1) under these assumptions is therefore bounded by the MSE of the finite Wiener filter, which is

$$e_{\text{ind}}^2 = E\left\{\left|d[n] - \hat{d}_{\text{ind}}[n]\right|^2\right\} = E\left\{|d[n]|^2\right\} - \left(\boldsymbol{R}^{-1}\vec{\boldsymbol{P}}\right)^H\vec{\boldsymbol{P}}. \quad (4)$$

IV. Optimal Estimator

The "independence assumptions" are very restrictive, exempting most signals found in communications systems. All the cases where the LMS filter outperformed the finite Wiener filter used signals which necessarily violated the "independence assumptions." Since our primary concern is bounding, not modeling, the performance of the LMS filter, we will now derive a bound using a set of assumptions which do not include the "independence assumptions" and which are valid for many of the signals used in the reported cases where performance surpasses that of the finite Wiener filter.

Once again writing (3), we have for the optimal estimator

$$\hat{d}_{\text{opt}}[n] = E\{d[n]|u[n], u[n-1], \cdots, u[-\infty], d[n-1], \cdots, d[-\infty]\}.$$

Eliminating the first two constraints and using only an expansion of independence Assumption 3), the mutually Gaussian assumption, to include the entirety of both processes, the optimal estimator is a linear estimate and is given by

$$\hat{d}_{\text{opt}}[n] = \sum_{i=-\infty}^{n} a_i u[i] + \sum_{i=-\infty}^{n-1} b_i d[i]. \quad (5)$$

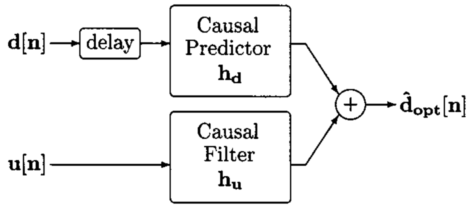

Fig. 3. Optimal estimator.

Note that (5) is a function of all the past and present reference data, and of all past samples of the desired signal. This equation can be rewritten as the output to the system shown in Fig. 3

$$\hat{d}_{\rm opt}[n] = \sum_{i=-\infty}^{n} h_u[n-i]u[i] + \sum_{i=-\infty}^{n-1} h_d[n-1-i]d[i] \quad (6)$$

where the impulse responses of the causal linear filter and causal linear predictor are given as $h_u[n]$ and $h_d[n]$, respectively.

Several definitions will now be made which will be used throughout the rest of this correspondence.

Definition 1: For $\delta > 0$, let $L_\delta(d\lambda)$ be the set of all functions $\{f(\lambda)\}$ such that

$$\int_{-\pi}^{\pi} |f(\lambda)|^\delta \, d\lambda < \infty.$$

Definition 2: For $\delta > 0$, let $\boldsymbol{L}_\delta$ be the set of all matrix functions $\{\boldsymbol{F}(\lambda)\}$ such that for every element of the matrix, $f_{i,j}(\lambda)$,

$$f_{i,j}(\lambda) \in L_\delta(d\lambda).$$

Definition 3: For $\delta > 0$, let l_δ be the set of all functions on the integer set $\{g[n]\}$ such that

$$\sum_{n=-\infty}^{\infty} |g[n]|^\delta < \infty.$$

The form of the optimal estimator is known from (6); solving for it requires finding the causal filters, $h_d[n]$ and $h_u[n]$, which result in the minimum MSE. To do this, the following assumptions will be made of the processes $\{u[n], -\infty < n < \infty\}$ and $\{d[n], -\infty < n < \infty\}$ and the filters $h_d[n]$, $h_u[n]$.

i) $u[n]$ and $d[n]$ are second-order processes:

$$E\left\{|u[n]|^2\right\} < \infty, \qquad \text{for all } n = 0, \pm 1, \cdots,$$
$$E\left\{|d[n]|^2\right\} < \infty, \qquad \text{for all } n = 0, \pm 1, \cdots,.$$

ii) $u[n]$ and $d[n]$ are zero mean:

$$E\{u[n]\} = m_u = 0 \qquad E\{d[n]\} = m_d = 0.$$

iii) The covariance functions depend only on the integer shifts k:

$$E\{(u[n+k]-m_u)(u[n]-m_u)^*\} = c_{uu}[k]$$
$$E\{(d[n+k]-m_d)(d[n]-m_d)^*\} = c_{dd}[k]$$
$$E\{(d[n+k]-m_d)(u[n]-m_u)^*\} = c_{du}[k].$$

iv) The covariance functions are absolutely summable:

$$c_{dd}[k] \in l_1 \qquad c_{uu}[k] \in l_1 \qquad c_{du}[k] \in l_1.$$

v) The spectral density matrix $\boldsymbol{\Phi}(\lambda)$ is nonnegative Hermitian with

$$\log|\det(\boldsymbol{\Phi}(\lambda))| \in L_1.$$

vi) h_d and h_u are stable filters:

$$h_d[k] \in l_1 \qquad h_u[k] \in l_1.$$

The optimal estimator is given by the conditional mean which is a projection operator. Let us first make use of the projection theorem [7, pp. 48–54], which states:

Theorem 1 (Projection Theorem): If M is a closed subspace of the Hilbert space H and $x \in H$, then there is a unique element $\hat{x} \in M$ such that

$$\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$$

and that

$$(x - \hat{x}) \in M^\perp.$$

The set of all second-order random variables defined on the probability space (Ω, F, P) forms a Hilbert space $L_2(dP)$ [7, pp. 46–48]. A closed linear subspace of $L_2(dP)$ is formed by the discrete-time convolution of stable filters acting on elements of $L_2(dP)$. Assumptions i) and vi) are therefore sufficient to invoke this theorem. Using the property that the error of the optimal estimator is orthogonal to the subspace, we can write

$$E\left\{\left(d[n] - \hat{d}_{\rm opt}[n]\right) \cdot (u[n], \cdots, u[-\infty], d[n-1], \cdots, d[-\infty])^*\right\} = \vec{0}.$$

This can be separated into two equations which must both be satisfied:

$$E\left\{\left(d[n] - \hat{d}_{\rm opt}[n]\right)(u[j])^*\right\} = 0, \qquad j \le n \quad (7)$$

and

$$E\left\{\left(d[n] - \hat{d}_{\rm opt}[n]\right)(d[j-1])^*\right\} = 0, \qquad j \le n. \quad (8)$$

To solve for the filters h_d and h_u, we represent the processes and filters in terms of their spectral representations. The spectral representation theorem [7, pp. 143–150] states:

Theorem 2 (Spectral Representation Theorem): For any wide-sense stationary second-order process $x[n]$, with spectral distribution function $S(\lambda)$, there exists an orthogonal increment process, $\xi(\lambda)$, such that

$$E\left\{|\xi(\lambda) - \xi(-\pi)|^2\right\} = S(\lambda), \qquad -\pi \le \lambda \le \pi \quad (9)$$

and

$$x[n] = \int_{-\pi}^{\pi} e^{in\lambda} \, d\xi(\lambda), \qquad \text{with probability one.} \quad (10)$$

Assumptions i)–iii) satisfy the conditions necessary for the spectral representation theorem. Using (10), we can write $d[n]$ and $u[n]$ in terms of their spectral representations:

$$d[n] = \int_{-\pi}^{\pi} e^{in\lambda} \, d\xi_d(\lambda) \quad (11)$$

and

$$u[n] = \int_{-\pi}^{\pi} e^{in\lambda} \, d\xi_u(\lambda). \quad (12)$$

From the spectral representation theorem, the output process resulting from the linear time-invariant filtering of a second-order wide-sense-stationary input process may be written in terms of the spectral representation of the input process and the Fourier transform of the filter $H(\lambda)$, provided that the filter resides in $L_2(dS(\lambda))$, where $S(\lambda)$ is the power spectral distribution of the input process [7, pp. 152–156]. To write $\hat{d}_{\rm opt}[n]$ in terms of the spectral representations $\xi_d(\lambda)$ and $\xi_u(\lambda)$, we therefore require that $\{d[n]\}$ and $\{u[n]\}$ be jointly wide-sense stationary and that the Fourier transforms of the filters $H_d(\lambda)$ and $H_u(\lambda)$ reside in $L_2(dS_d(\lambda))$ and $L_2(dS_u(\lambda))$, respectively. Since, by Assumption vi), we require the stricter condition that the filters $h_d[n]$ and $h_u[n]$ each reside in l_1, we satisfy this requirement and can therefore write

$$\hat{d}_{\rm opt}[n] = \int_{-\pi}^{\pi} e^{-i\lambda} H_d(\lambda)\, e^{in\lambda}\, d\xi_d(\lambda) + \int_{-\pi}^{\pi} H_u(\lambda)\, e^{in\lambda}\, d\xi_u(\lambda). \tag{13}$$

Applying (11)–(13) to the projection theorem, (7) and (8), gives

$$E\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(n\lambda - j\gamma)}\, [d\xi_d(\lambda)\, d\xi_u^*(\gamma) - H_d(\lambda)e^{-i\lambda}\, d\xi_d(\lambda)\, d\xi_u^*(\gamma) - H_u(\lambda)\, d\xi_u(\lambda)\, d\xi_u^*(\gamma)]\right\} = 0, \qquad j \le n \tag{14}$$

and

$$E\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(n\lambda - j\gamma)}\, \left[e^{i\lambda}\, d\xi_d(\lambda)\, d\xi_d^*(\gamma) - H_d(\lambda) d\xi_d(\lambda)\, d\xi_d^*(\gamma) - e^{i\lambda}\, H_u(\lambda)\, d\xi_u(\lambda)\, d\xi_d^*(\gamma)\right]\right\} = 0, \qquad j \le n. \tag{15}$$

Since these integrals are defined in the mean-square sense, the expectation operation can be pulled inside the integrations. Swapping integration and expectation in (14) and (15) results in

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(n\lambda - j\gamma)}\, (E\{d\xi_d(\lambda) d\xi_u^*(\gamma)\} - H_d(\lambda)e^{-i\lambda} E\{d\xi_d(\lambda) d\xi_u^*(\gamma)\} - H_u(\lambda) E\{d\xi_u(\lambda) d\xi_u^*(\gamma)\}) = 0, \qquad j \le n \tag{16}$$

and

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(n\lambda - j\gamma)}\, \Big(e^{i\lambda} E\{d\xi_d(\lambda) d\xi_d^*(\gamma)\} - H_d(\lambda) E\{d\xi_d(\lambda) d\xi_d^*(\gamma)\} - e^{i\lambda} H_u(\lambda) E\{d\xi_u(\lambda) d\xi_d^*(\gamma)\}\Big) = 0, \quad j \le n. \tag{17}$$

The orthogonal increment property of the spectral representation, $\xi(\lambda)$, along with (9) from the spectral representation theorem, yields [7, pp. 138–140]

$$E\{d\xi(\lambda) d\xi^*(\gamma)\} = \delta_{\lambda,\gamma}\, dS(\lambda),$$

where $\delta_{\lambda,\gamma}$ is the Kronecker delta function. This can be extended to multiple processes $x[n]$ and $y[n]$, provided the processes are jointly wide-sense-stationary [7, pp. 454–459], giving

$$E\left\{d\xi_x(\lambda) d\xi_y^*(\gamma)\right\} = \delta_{\lambda,\gamma} dS_{xy}(\lambda).$$

Applying these to (16) and (17), we are able to write them as single integrals over the spectral distribution functions, namely,

$$\int_{-\pi}^{\pi} e^{i(n-j)\lambda} \Big(dS_{du}(\lambda) - H_d(\lambda)e^{-i\lambda} dS_{du}(\lambda) - H_u(\lambda)\, dS_{uu}(\lambda)\Big) = 0, \qquad j \le n \tag{18}$$

and

$$\int_{-\pi}^{\pi} e^{i(n-j)\lambda} \Big(e^{i\lambda}\, dS_{dd}(\lambda) - H_d(\lambda)\, dS_{dd}(\lambda) - e^{i\lambda} H_u(\lambda) dS_{ud}(\lambda)\Big) = 0, \qquad j \le n. \tag{19}$$

To write (18) and (19) in terms of the processes' spectral density functions, we need the spectral density functions to exist and to be bounded and continuous. This can be related to constraints on the processes' covariance functions through Herglotz's theorem [7, pp. 117–122], which states: The spectral distribution function is related to the covariance function of a process by

$$c[n] = \int_{-\pi}^{\pi} e^{in\lambda}\, dS(\lambda), \qquad \text{for all } n = 0, \pm 1, \cdots,. \tag{20}$$

From (20) it can be shown that if the covariance function of a process is absolutely summable, then its spectral density function exists and is bounded and continuous [7, pp. 117–122]. The spectral density function can then be obtained from the spectral distribution function by

$$s(\lambda) = \frac{dS(\lambda)}{d\lambda}.$$

Assumption iv) thus allows us to write (18) and (20) in terms of the spectral density functions:

$$\int_{-\pi}^{\pi} e^{i(n-j)\lambda} \Big(s_{du}(\lambda) d\lambda - H_d(\lambda)e^{-i\lambda} s_{du}(\lambda) d\lambda - H_u(\lambda) s_{uu}(\lambda) d\lambda\Big) = 0, \qquad j \le n \tag{21}$$

and

$$\int_{-\pi}^{\pi} e^{i(n-j)\lambda} \Big(e^{i\lambda} s_{dd}(\lambda) d\lambda - H_d(\lambda) s_{dd}(\lambda) d\lambda - e^{i\lambda} H_u(\lambda) s_{ud}(\lambda) d\lambda\Big) = 0, \qquad j \le n. \tag{22}$$

Writing (21) and (22) as a single matrix equation, we get

$$\int_{-\pi}^{\pi} e^{im\lambda} \Big[\vec{\mathbf{\Gamma}}(\lambda) - \mathbf{\Phi}(\lambda)\vec{\boldsymbol{H}}(\lambda)\Big]\, d\lambda = 0, \qquad \forall m \ge 0 \tag{23}$$

where

$$\mathbf{\Phi}(\lambda) = \begin{bmatrix} s_{uu}(\lambda) & s_{du}(\lambda)e^{-i\lambda} \\ s_{ud}(\lambda)e^{i\lambda} & s_{dd}(\lambda) \end{bmatrix}$$

$$\vec{\mathbf{\Gamma}}(\lambda) = \Big[s_{du}(\lambda),\, s_{dd}(\lambda)e^{i\lambda}\Big]^T$$

and

$$\vec{\boldsymbol{H}}(\lambda) = [H_u(\lambda),\, H_d(\lambda)]^T.$$

It should be noted that in the case of a one-step predictor the constraints (21) and (22) are identical, since the reference signal is a single sample delayed version of the desired signal. For this degenerate case, we use only one constraint and the scalar versions of the matrix equations which follow.

In addition to satisfying (23), $\vec{\boldsymbol{H}}(\lambda)$ must also be causal. The causality constraint prevents us from simply using matrix inversion of the spectral density matrix, $\mathbf{\Phi}(\lambda)$, to solve for $\vec{\boldsymbol{H}}(\lambda)$. Instead, we desire to break $\mathbf{\Phi}(\lambda)$ into its causal and anticausal portions, thus allowing us to solve for the causal $\vec{\boldsymbol{H}}(\lambda)$. This is known as the spectral factorization of $\mathbf{\Phi}(\lambda)$, and is not universally possible. Wiener and Masani in [11] provide the following existence theorem.

Theorem 3 (Wiener and Masani): Given a nonnegative hermitian matrix-valued function $\boldsymbol{F}$ on the set of complex numbers C, such that $\boldsymbol{F} \in \boldsymbol{L}_1$ and $\log(\det(\boldsymbol{F})) \in L_1(d\lambda)$, there exists a function $\mathbf{X} \in \boldsymbol{L}_2$ on C, the nth Fourier coefficient of which vanishes for $n < 0$, such that

$$\boldsymbol{F}(e^{i\lambda}) = \boldsymbol{X}(e^{i\lambda})\boldsymbol{X}^H(e^{i\lambda}) \qquad \text{a.e.}$$

Assumptions iv) and v) satisfy the above theorem for the spectral density matrix. Note that Wiener and Masani defined the nth Fourier coefficient of $\boldsymbol{X}$ as

$$\boldsymbol{C}[n] = \int_{-\pi}^{\pi} e^{-in\lambda} \boldsymbol{X}(e^{i\lambda})\, d\lambda$$

whereas we have defined the nth Fourier coefficients implicitly through Herglotz's theorem as

$$\boldsymbol{C}[n] = \int_{-\pi}^{\pi} e^{in\lambda} \boldsymbol{X}(e^{i\lambda})\, d\lambda.$$

For us $\boldsymbol{X}(e^{i\lambda})$ thus represents the anticausal factor. We will therefore write the spectral density matrix as the product of a matrix $\boldsymbol{\Psi}(\lambda)$ and its hermitian transpose, where $\boldsymbol{\Psi}(\lambda)$ represents the causal part and $\boldsymbol{\Psi}^H(\lambda)$ represents the anticausal part:

$$\boldsymbol{\Phi}(\lambda) = \boldsymbol{\Psi}^H(\lambda)\boldsymbol{\Psi}(\lambda).$$

Note that many methods, both analytic and numeric, exist to perform such matrix factorizations; the reader is referred to Kucera [12] for an overview of several methods. The solution to the multidimensional filtering/prediction problem can then be solved, using a method demonstrated by Wong [13], as follows:

$$\int_{-\pi}^{\pi} e^{im\lambda} \left[\vec{\boldsymbol{\Gamma}}(\lambda) - \boldsymbol{\Phi}(\lambda)\vec{\boldsymbol{H}}(\lambda)\right] d\lambda = \vec{\boldsymbol{f}}(m)$$

where

$$\vec{\boldsymbol{f}}(m) = \vec{\boldsymbol{0}}, \qquad m \geq 0.$$

Writing $\vec{\boldsymbol{f}}(m)$ in terms of its Fourier transform allows us to write

$$\left[\vec{\boldsymbol{\Gamma}}(\lambda) - \boldsymbol{\Phi}(\lambda)\vec{\boldsymbol{H}}(\lambda)\right] = \vec{\boldsymbol{F}}(\lambda).$$

Premultiplying both sides by $[\boldsymbol{\Psi}^H(\lambda)]^{-1}$, we have

$$\left[\boldsymbol{\Psi}^H(\lambda)\right]^{-1} \vec{\boldsymbol{\Gamma}}(\lambda) - \boldsymbol{\Psi}(\lambda)\vec{\boldsymbol{H}}(\lambda) = \left[\boldsymbol{\Psi}^H(\lambda)\right]^{-1} \boldsymbol{F}(\lambda).$$

Since $[\boldsymbol{\Psi}^H(\lambda)]^{-1}\boldsymbol{F}(\lambda)$ is anticausal and $\boldsymbol{\Psi}(\lambda)\vec{\boldsymbol{H}}(\lambda)$ is causal, $\boldsymbol{\Psi}(\lambda)\vec{\boldsymbol{H}}(\lambda)$ must equal the causal part of $[\boldsymbol{\Psi}^H(\lambda)]^{-1}\vec{\boldsymbol{\Gamma}}(\lambda)$. The optimal filters are then given by

$$\vec{\boldsymbol{H}}(\lambda) = [\boldsymbol{\Psi}(\lambda)]^{-1} \left\{ \left[\boldsymbol{\Psi}^H(\lambda)\right]^{-1} \vec{\boldsymbol{\Gamma}}(\lambda) \right\}^{+}$$

where $\{\,\}^+$ refers to the causal portion. Note that using this method will give filter solutions $H_u(\lambda) \in L_2(dS_u(\lambda))$ and $H_d(\lambda) \in L_2(dS_d(\lambda))$. This results from our use of the spectral representation theorem in describing the estimate $\hat{d}_{opt}[n]$ in (13). The theorem requires only that the processes be second-order and wide-sense stationary; this restricts the filters to lie in $L_2(dS_u(\lambda))$ and $L_2(dS_d(\lambda))$, respectively. The structure of our estimate, however, requires the stricter condition that the filters $h_u[n]$ and $h_d[n]$ reside in l_1. This method therefore does not guarantee us a solution which matches our assumptions; however, such a case has not been encountered.

The MSE of the optimal estimator, and thus a bound on the LMS estimator's performance, is

$$e_{\text{opt}}^2 = E\left\{\left|d[n] - \hat{d}_{\text{opt}}[n]\right|^2\right\}. \tag{24}$$

Substituting (13) in for $\hat{d}_{\text{opt}}[n]$ allows us to write (24) in terms of the filters $H_d(\lambda)$ and $H_u(\lambda)$

$$E\left\{\left|d[n] - \int_{-\pi}^{\pi} e^{-i\lambda} H_d(\lambda) e^{in\lambda}\, d\xi_d(\lambda) - \int_{-\pi}^{\pi} H_u(\lambda) e^{in\lambda}\, d\xi_u(\lambda)\right|^2\right\}. \tag{25}$$

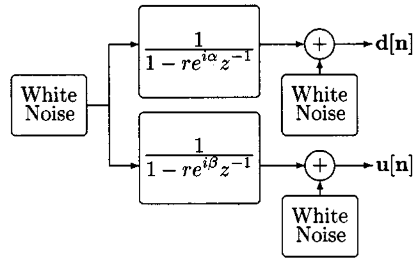

Fig. 4. Simulation system.

As previously, by expanding (25) and swapping integration and expectation, we are able to write this in terms of the processes' spectral density functions, as

$$\begin{aligned} e_{\text{opt}}^2 = \int_{-\pi}^{\pi} s_{dd}(\lambda) d\lambda - \int_{-\pi}^{\pi} |H_d(\lambda)|^2\, s_{dd}(\lambda)\, d\lambda \\ - 2\Re\left(\int_{-\pi}^{\pi} e^{-i\lambda} H_d(\lambda) H_u^*(\lambda) s_{du}(\lambda)\, d\lambda\right) \\ - \int_{-\pi}^{\pi} |H_u(\lambda)|^2\, s_{uu}(\lambda)\, d\lambda. \end{aligned} \tag{26}$$

Equation (26) is the performance of the optimal estimator under these assumptions and, as such, it is a bound on the performance of all other estimators. Therefore, this provides us with a bound on the performance of the LMS estimator under a mild set of assumptions which does not exclude data contributions available to the estimator.

V. Results

Now that we have bounded the performance of the LMS estimator, (26), for the class of signals meeting Assumptions i)–v), we wish to demonstrate this bound for scenarios where the LMS estimator outperforms the finite Wiener filter. To do this, we must produce reference and desired signals which meet Assumptions i)–v) and, when operated on by an LMS estimator, result in that estimator's performance exceeding that of the finite Wiener filter.

The stable autoregressive (AR) processes generated by the system in Fig. 4 satisfy Assumptions i)–v). The desired and reference signals are first-order AR processes generated from the same white Gaussian noise source with an independent Gaussian noise component added to each signal.

For the LMS adaptive filter to outperform the finite Wiener filter, it is necessary to generate scenarios where the optimal estimator, and thus the LMS filter, have significant contributions from data not available to the finite Wiener filter. This is not to say that this is sufficient to generate such scenarios, just that it is necessary. The filters of the optimal estimator, $H_d(\lambda)$ and $H_u(\lambda)$, are directly dependent on the auto- and crosscorrelations of the desired and reference processes. When the optimal estimator has significant contributions from data outside the length of the finite Wiener filter, we would expect the auto- and crosscorrelations to have significant magnitude outside the range of the finite Wiener filter. While the converse is not necessarily true, a trivial example being where the reference and desired processes have long autocorrelations but are identical and result in the trivial estimator of $\hat{d}[n] = u[n]$, this does direct us toward the types of processes which might generate this behavior. As the pole of an AR process is moved closer to the unit circle, the magnitude and significant length of the process's autocorrelation function is increased. Therefore, in order to have the system in Fig. 4 generate scenarios where the LMS performance exceeds that of the finite Wiener filter, the poles of the AR processes are restricted to be close to the unit circle.

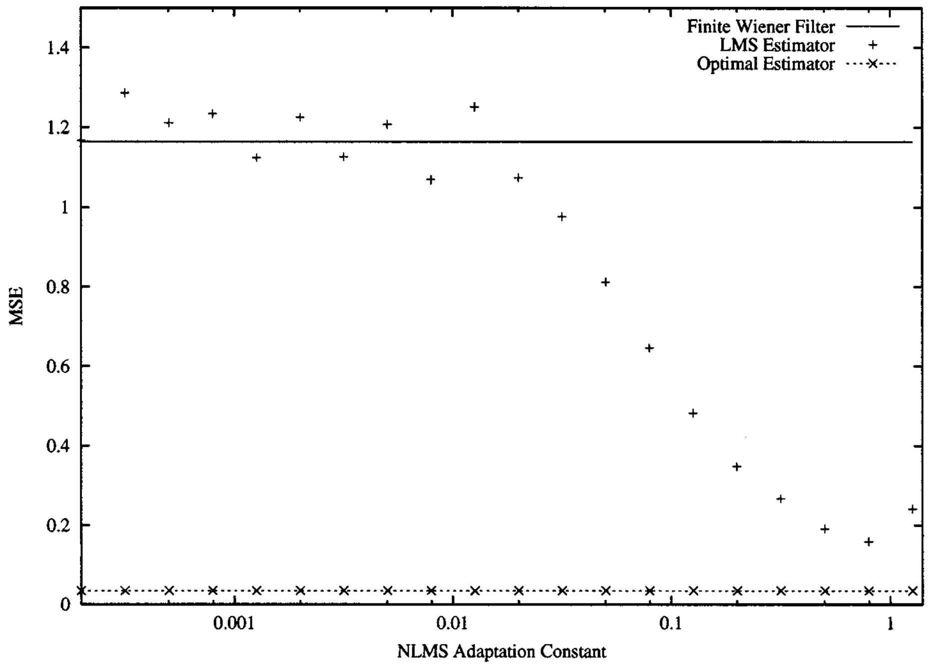

Fig. 5. MSE as a function of the NLMS adaptation constant, μ, with $\mathrm{SNR}_d = 20$ dB, $\mathrm{SNR}_u = 20$ dB, $r = 0.99$, $\alpha - \beta = 3.6^\circ$, and $L = 25$.

The MSE performance of the LMS estimator was evaluated through Monte Carlo simulations. Instead of the standard LMS weight update algorithm, the normalized least mean-square algorithm (NLMS) was used. The weight update algorithm for NLMS is given in [10, pp. 432–437] as

$$\vec{w}[n] = \vec{w}[n-1] + \frac{\mu e^*[n-1]\vec{u}_L[n-1]}{\vec{u}_L^H[n-1]\vec{u}_L[n-1]}.$$

This was done because the NLMS algorithm was stable over relatively larger values of the adaptation constant, where the performance gain was greatest. The finite Wiener filter and optimal estimator performances were evaluated numerically. The matrix spectral factorization required in evaluating the optimal estimator was performed using the analytical method developed by Davis [14], modified for a complex discrete system. Plots of the MSE performance of the LMS estimator, optimal estimator, and finite Wiener filter are shown under different combinations of pole locations and signal-to-noise ratios in order to compare the performance of the LMS estimator to both an absolute bound on its performance as well as to the bound obtained through the "independence assumptions."

Fig. 5 is a graph of the MSE as a function of the adaptation constant μ. For this simulation, the AR poles were held constant with radii $r = 0.99$ and angle separation 3.6°. The signal-to-noise ratio on the desired channel, SNR_d, was 20 dB, while the reference channel signal-to-noise ratio, SNR_u, was also 20 dB. The filter length L was 25 taps. From the plot, notice that the LMS estimator does indeed outperform the finite Wiener filter. Also, note that the LMS estimator performance comes close to achieving that of the optimal estimator as μ increases to an optimal value. Having a decreasing MSE as μ increases is contrary to the conventional wisdom that has resulted from the "independence assumptions" model. This can be explained by noting that μ regulates the amount of error fed back into the LMS estimator; by increasing the amount of error being fed back, one is increasing the contributions of the data from the past values of desired and reference data, which are unavailable to the finite Wiener filter.

Fig. 6 is a graph of MSE as a function of the pole radii r, where for each value of r the LMS estimator MSE values were taken over a range of adaptation constants; the one which resulted in the smallest MSE was used. The pole angle separation, $\alpha - \beta$, was held constant at 3.6°, with signal-to-noise ratios of $\mathrm{SNR}_d = 20$ dB and $\mathrm{SNR}_u = 20$ dB and a filter length of $L = 25$. As a pole is moved closer to the unit circle, the power of the AR process is increased. To maintain constant signal-to-noise ratios, the AR process generator noise power was decreased as the poles were moved closer to the unit circle, maintaining a constant power for the AR processes. The performance of the optimal estimator is almost invariant to the pole locations, and is instead dependent on the signal-to-noise ratios in the reference and desired channels. Also, note that it is not until the poles begin to approach the unit circle that the optimal estimator and the finite Wiener filter's performance begin to diverge. The LMS estimator's performance was dependent on the pole locations, but not to the same degree as the finite Wiener filter; its MSE rose more slowly than that of the finite Wiener filter as the poles approached the unit circle.

Fig. 7 is a graph of MSE as a function of the pole angle difference, and again the LMS estimator's performance was evaluated over a range of adaptations constants, the one with the lowest MSE being used. The pole radii were held constant at $r = 0.99$, with signal-to-noise ratios of $\mathrm{SNR}_d = 20$ dB and $\mathrm{SNR}_u = 20$ dB, and a filter length of $L = 25$. The smaller the angle separation, the greater crosscorrelation between the reference and the desired process. This, along with the long autocorrelation sequences resulting from the poles' proximity to the unit circle, allows the LMS estimator to take advantage of correlated data outside of the range of the finite Wiener filter, resulting in lower MSE for the LMS estimator. Thus as the pole angle separation decreases, the disparity between the LMS estimator and the finite Wiener filter increases, until the angle separation nears zero. At these points, the high crosscorrelation between $u[n]$ and $d[n]$ dominates the terms in the optimal estimator, allowing for the finite Wiener filter to achieve similar performance.

Fig. 8 is a graph of the MSE as a function of the reference channel signal-to-noise ratio, SNR_u. The powers of the AR processes were held constant, with pole angle radii of $r = 0.99$ and angle separation $\alpha - \beta = 3.6^\circ$; only the reference noise power was varied. The LMS estimator's performance was evaluated at the optimum μ value. The filter

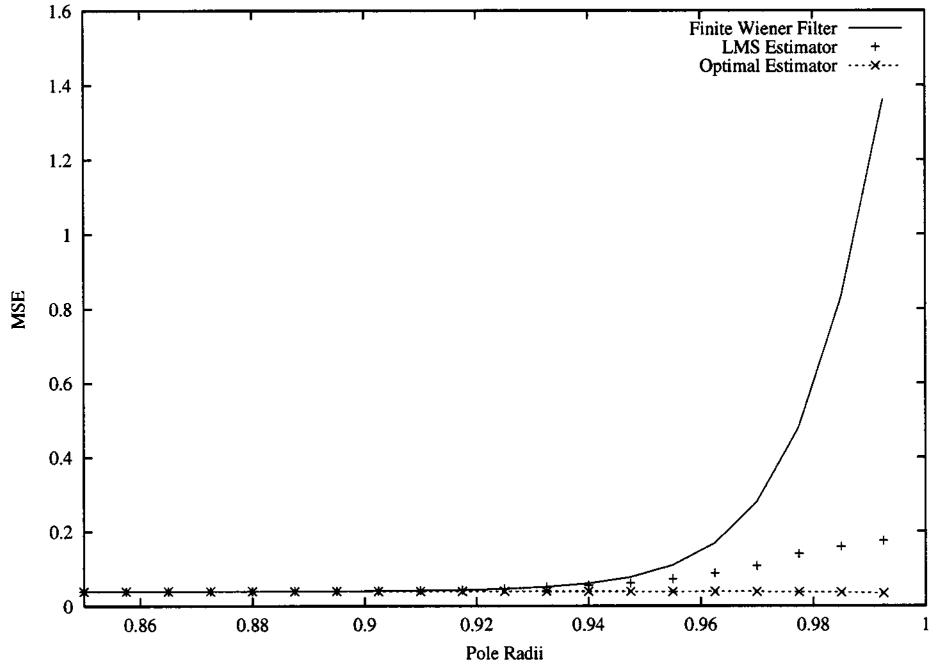

Fig. 6. MSE as a function of the pole radii r, with $\mathrm{SNR}_d = 20$ dB, $\mathrm{SNR}_u = 20$ dB, $\alpha - \beta = 3.6°$, $\mu = \mu_{\mathrm{opt}}$, and $L = 25$.

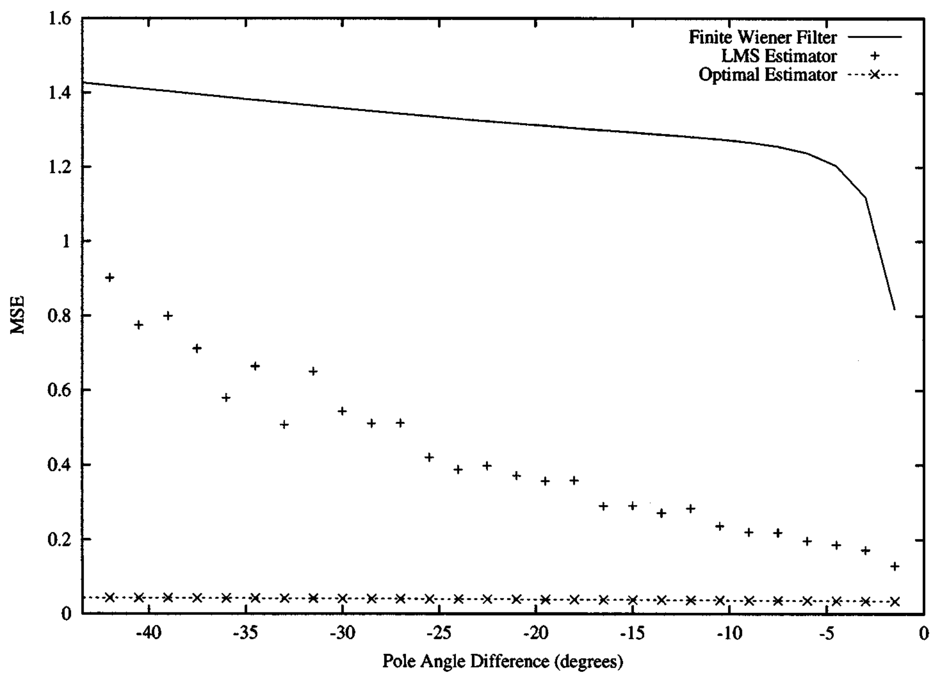

Fig. 7. MSE as a function of the pole angle difference, $\alpha - \beta$, with $\mathrm{SNR}_d = 20$ dB, $\mathrm{SNR}_u = 20$ dB, $r = 0.99$, $\mu = \mu_{\mathrm{opt}}$, and $L = 25$.

length of the LMS estimator was $L = 25$. The LMS estimator (2) is dominated by terms from the reference channel; this dependence is exemplified by the degradation of the LMS estimator's performance for low SNR_u values. Fig. 9 is the complementary graph to Fig. 8, where now the signal-to-noise ratio of the desired channel, SNR_d, is varied and the SNR_u is held constant at 20 dB. Comparing this to Fig. 8, the performance of the LMS estimator is less sensitive to changes on the desired channel SNR. Given that the optimal performance is relatively constant, when the SNR on either channel becomes low it uses more of the signal from the other channel to form the estimate. Curves representing the performance of a causal Wiener predictor, operating on the past samples of the desired channel, and of a causal Wiener filter, operating on the present and past samples of the reference channel, are included in Figs. 8 and 9 to corroborate this point. Comparing, in Fig. 8, the optimal estimator to that of both the causal Wiener filter and the Wiener predictor operating at a low reference channel SNR, the optimal estimator is clearly relying more heavily on the desired channel data, given that the optimal estimator's performance nearly matches that of the causal Weiner predictor and far exceeds that of the causal Wiener filter. The same can be shown for the converse case where the desired channel SNR is low, where in this case the optimal estimator's performance more closely matches that of the causal Wiener filter which uses

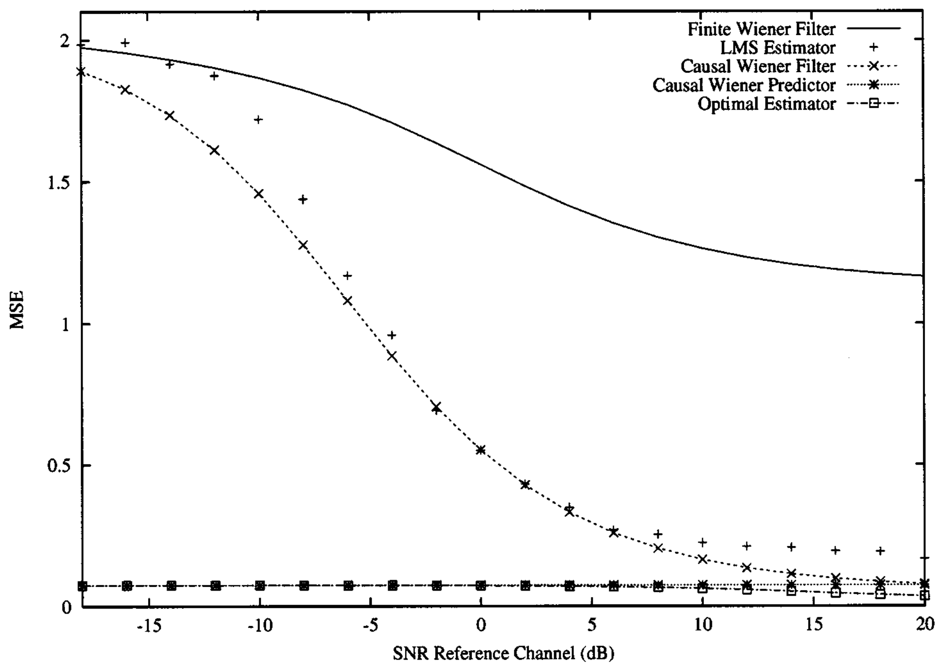

Fig. 8. MSE as a function of the signal-to-noise ratio on the reference channel, SNR_u, with $\mathrm{SNR}_d = 20$ dB, $r = 0.99$, $\alpha - \beta = 3.6°$, $\mu = \mu_{\mathrm{opt}}$, and $L = 25$.

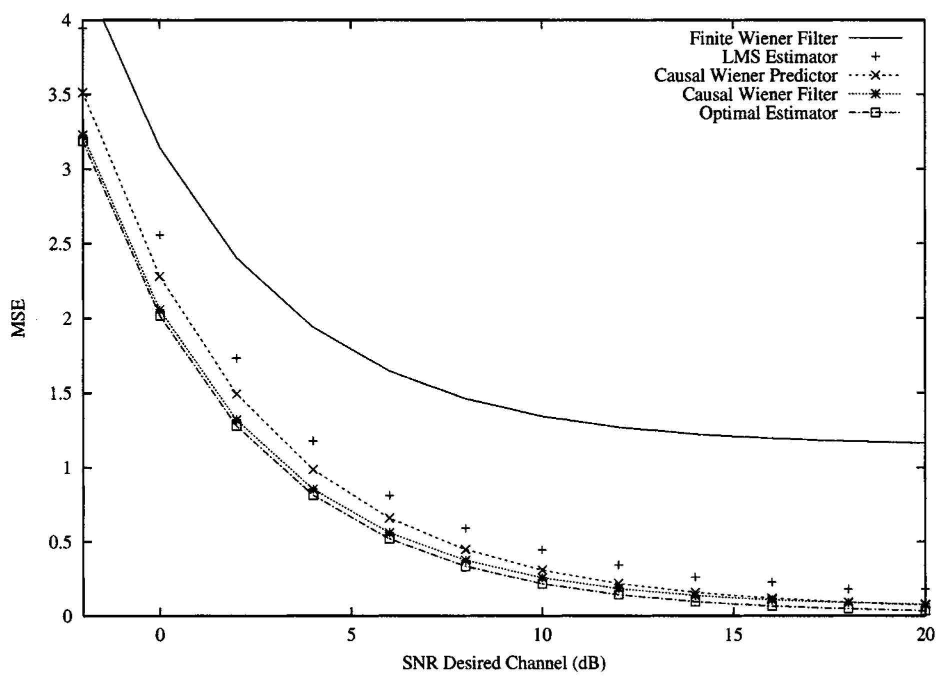

Fig. 9. MSE as a function of the signal-to-noise ratio on the desired channel, SNR_d, with $\mathrm{SNR}_u = 20$ dB, $r = 0.99$, $\alpha - \beta = 3.6°$, $\mu = \mu_{\mathrm{opt}}$, and $L = 25$.

only data from the reference channel. In contrast, the LMS estimator is tightly tied to the reference channel, so degradation on this channel can lead to LMS estimator performance degradation equal to that of the finite Wiener filter. This dependence is exemplified by comparing the LMS estimator's performance with that of the causal Wiener filter and the Wiener predictor and noting that its performance more closely resembles that of the causal Wiener filter. This is not to imply that the LMS only makes use of the reference channel; as shown in Section IV it, in fact, uses data from both channels and thus its performance is only bounded by that of the optimal estimator(26). This is demonstrated in Fig. 8 where the LMS estimator, a nonlinear estimator formed from a simple feedback structure, is shown to match the performance of the causal Wiener filter operating on all past and present samples of the reference channel.

VI. Conclusion

In conclusion, we bounded the performance of the LMS estimator for a class of signals without using the "independence assumptions." Cases where the LMS estimator outperforms the finite Wiener filter which, again, arises from the "independence assumptions" bound, were then generated, and the performance of the LMS estimator was compared to the bound. The LMS estimator was found to greatly outperform the finite Wiener filter in these cases, and, in some cases, resembled that of the optimal estimator. This behavior can be attributed to the fact that the LMS estimator uses information not available to the finite Wiener filter. These data include not only all values of the reference data, but also all past values of the desired signal. Note that even though the LMS estimator uses past values of the desired signal, it is more dependent on the values of the reference data; this was demonstrated by the sensitivity to SNR changes on the reference channel. Finally, while the effect demonstrated in this correspondence occurs only under severe violation of the "independence assumptions," it does have applications in both interference suppression and noise cancellation. Both [3] and [4] have demonstrated the usefulness of this property when the LMS estimator is used in adaptive equalizers to suppress a narrowband interferer. As for noise cancellation, the simulation results presented in this correspondence could be easily reworked to represent the LMS estimator when it is used to cancel a Doppler shifted narrowband signal.

REFERENCES

[1] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong Jr, and R. Goodin, "Adaptive noise canceling: principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.

[2] B. Widrow, J. McCool, M. Larimore, and R. Johnson Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[3] R. North, R. Axford, and J. Zeidler, "The performance of adaptive equalization for digital communications systems corrupted by interference," in *27th ASILOMAR Conf. Signals, Systems and Computers*, 1993, pp. 1548–1553.

[4] M. Reuter and J. R. Zeidler, "Nonlinear effects in LMS adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 47, pp. 1570–1579, June 1999.

[5] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 43, pp. 2863–2871, Dec. 1995.

[6] H. J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption," in *1995 Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-95)*, Detroit, MI, 1995, pp. 1404–1407.

[7] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer, 1991.

[8] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 722–736, June 1981.

[9] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 222–230, Feb. 1985.

[10] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[11] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes—I: The regularity condition," *Acta Math.*, pp. 111–150, 1957.

[12] V. Kucera, "Factorization of rational spectral matrices: A survey of methods," in *IEE Int. Conf. Control '91*, vol. 2, Edinburgh, Scotland, U.K., Mar. 1991, pp. 1047–1078.

[13] E. Wong, "On the multidimensional prediction and filtering problem and the factorization of spectral matrices," *J. Franklin Inst.*, vol. 272, no. 2, pp. 87–99, Aug. 1961.

[14] M. C. Davis, "Factoring the spectral matrix," *IEEE Trans. Automat. Contr.*, vol. AC-8, pp. 296–305, Oct. 1963.

A Minimax Robust Decoding Algorithm

Lei Wei, *Senior Member, IEEE*, Zheng Li, *Student Member, IEEE*, Matthew R. James, *Member, IEEE*, and Ian R. Petersen, *Fellow, IEEE*

***Abstract*—In this correspondence we study the decoding problem in an uncertain noise environment. If the receiver knows the noise probability density function (pdf) at each time slot or its *a priori* probability, the standard Viterbi algorithm (VA) or the *a posteriori* probability (APP) algorithm can achieve optimal performance. However, if the actual noise distribution differs from the noise model used to design the receiver, there can be significant performance degradation due to the model mismatch. The minimax concept is used to minimize the worst possible error performance over a family of possible channel noise pdf's. We show that the optimal robust scheme is difficult to derive; therefore, alternative, practically feasible, robust decoding schemes are presented and implemented on VA decoder and two-way APP decoder. Performance analysis and numerical results show our robust decoders have a performance advantage over standard decoders in uncertain noise channels, with no or little computational overhead. Our robust decoding approach can also explain why for turbo decoding overestimating the noise variance gives better results than underestimating it.**

***Index Terms*—BCJR algorithm, impulsive noise, minimax, min-sum algorithm, robust signal processing, sum-product algorithm, turbo decoding, two-way APP algorithm, uncertain channel, Viterbi algorithm.**

I. INTRODUCTION

Over the last 30 years the Viterbi algorithm (**VA**) has been widely applied in digital communications, [1], [2]. Recently, the two-way *a posteriori* probability (**APP**) decoder [3]–[6] has also attracted a lot of attention, due to its application in turbo decoding [7]. Both the VA and two-way APP algorithms are special cases of min-sum and sum-product algorithms [9]. Some significant developments in understanding these two-way decoding algorithms have been reported in [8], [10]–[12]. In [13], Forney provided a detailed description of the two-way algorithms as well as a review of their rich history.

A. Decoding in Uncertain Noise Environment

In this correspondence, we aim to extend the two-way decoding algorithms to situations with unknown noise environments. As we know, if the probability density function (**pdf**) of the noise is known and constant over the decoding period, the conventional VA and APP algorithms are optimal; otherwise, a significant performance degradation may occur. In other words, we need to know the exact noise pdf of the channel to design the optimal standard decoder.

However, in practice, the pdf of the noise could change within a short time frame in an uncertain manner. This could be caused by either a natural phenomenon such as lighting or man-made noise such as automotive noise and power-line noise [17], [18]. The characteristic of one type of man-made noise is impulsive with a typical rate of 10–500 impulses/s, [17]. For a mobile phone with a data rate of 10 kbits/s, it could experience up to one impulse every 200 bits (roughly every speech packet will be affected). For an HF radio with a data rate of 1

Manuscript received December 9, 1998; revised January 26, 1999. The material in this correspondence was presented in part at the International Conference in Communications, Vancouver, BC, Canada, June 1999.

L. Wei, Z. Li, and M. R. James are with the Department of Engineering, Faculty of Engineering and Information Technology, Australian National University, Canberra, ACT 0200, Australia (e-mail: Lei.Wei@anu.edu.au).

I. R. Petersen is with the Department of Electrical Engineering, Australian Defence Force Academy, Canberra, ACT 2601, Australia.

Communicated by T. E. Fuja, Associate Editor At Large.

Publisher Item Identifier S 0018-9448(00)03098-4.