Non-Linear Effects in LMS Adaptive Filters

Michael Reuter  Kevin Quirk  James Zeidler  Laurence Milstein
SSC San Diego  UC San Diego  SSC San Diego  UC San Diego
San Diego, CA, USA  La Jolla, CA, USA  San Diego, CA, USA  La Jolla, CA, USA
reuter@spawar.navy.mil  kquirk@ucsd.edu  zeidler@spawar.navy.mil  milstein@ece.ucsd.edu

Abstract

Recent results demonstrate that adaptive filters implemented with the least mean square (LMS) algorithm can exhibit better mean square error (MSE) performance than the corresponding Wiener filter. We examine some conditions under which this can occur for implementations of an adaptive noise canceler and an adaptive equalizer. In particular, we demonstrate that because of the recursive LMS update equation, the LMS estimator is non-linear and uses much more information than that used by the corresponding Wiener filter. Under certain circumstances, this extra information can enhance the MSE performance of LMS over the Wiener filter. To quantify this effect, we use a transfer function approach to approximate the MSE of the LMS estimator. We also show that the LMS estimator is indeed bounded in MSE performance by a linear Wiener filter that explicitly uses the same information used in the LMS estimator.

1. Introduction

The minimum mean square error (MSE) performance of an adaptive filter implemented with the least mean square (LMS) algorithm is often assessed using the MSE of the corresponding Wiener filter. This approach is justified by the assumption that gradient noise on the LMS adaptive filter weights results in MSE that is greater than the Wiener MSE [1, p. 366]. Traditional approaches to analyzing LMS MSE model the gradient noise, or misadjustment component of the weights, as a stochastic vector process that is statistically independent of the current reference data vector [1, p. 393]. This approach results in an analytical expression of LMS MSE that is greater than the Wiener MSE and has agreed with experimental results for a variety of adaptive filter applications such as the adaptive noise canceler (ANC) [2], and the adaptive equalizer (AEQ) [3].

However, recently it has been reported that an AEQ implemented with LMS operating in the presence of narrowband interference can produce better probability of bit error and MSE performance than the Wiener filter [4, 5]. This effect has also been observed in an ANC in which the interference to be canceled is also narrowband [6]. The temporal correlation of the interference in both of these scenarios results in a severe violation of the independence assumption.

In this paper, we review the results presented in [5, 6] and extend the results by incorporating a more general ideal narrowband interference process into the AEQ scenario. We also apply the LMS analysis approach from [5] to the ANC problem presented in [6]. We support the theory with simulations and include some results from the recursive least-squares (RLS) estimator for purposes of comparison.

2. The LMS Estimator

The general problem is to estimate the desired discrete time signal $d(k)$ using past values $d(k-1), d(k-2), \ldots$, and samples from a reference signal $\{u(k)\}$. The LMS estimate $y(k)$ is given by

$$y(k) = w_h(k)u(k),$$  \hspace{1cm} (1)

where $w(k)$ is a vector of complex weights of length $L$, $u(k)$ is a vector comprised of $L$ samples of the reference signal $\{u(k)\}$, and $H$ denotes the Hermitian transpose. The weights are updated as

$$w(k+1) = w(k) + \mu(k)e^*(k)u(k),$$  \hspace{1cm} (2)

where $e(k) = d(k) - y(k)$, and $*$ denotes complex conjugation. The step size parameter is $\mu(k) = \mu$ for standard LMS and $\mu(k) = \mu/||u(k)||^2$ for normalized LMS (NLMS).

---

This work was supported by the Independent Research Program at SSC San Diego, and by the NSF Industry/University Cooperative Research Center on Integrated Circuits and Systems at UCSD.
Using the recursion of Eq. (2), the LMS estimator is a non-linear function of the data and can be written abstractly as
\[ y(k) = C_{\text{LMS}}(u(k) u(k-1) \ldots; d(k-1) d(k-2) \ldots), \]  
with MSE given by \( J_{\text{LMS}} = E[|e(k)|^2] \).

The optimum MSE estimate using this same information is given by
\[ C_o(u(k) u(k-1) \ldots; d(k-1) d(k-2) \ldots) = E[d(k)|u(k), u(k-1) \ldots; d(k-1) d(k-2) \ldots], \]  
where \( E[\cdot] \) denotes expectation taken over all random components. By using the independence assumptions given by [6]

**IA-1.** The composite variates \((d(k), u(k)), \ldots\) are statistically independent.

**IA-2.** \(d(k)\) is dependent on \(u(k)\),

the optimum MSE estimator is only a function of \(u(k)\) and can be written as
\[ C_o(u(k)) = E[d(k)|u(k)]. \]  
Making the additional assumption that \(\{d(k)\}\) and \(\{u(k)\}\) are jointly Gaussian distributed and jointly wide sense stationary (WSS), the optimum MSE estimator is linear and is given by the Wiener filter output
\[ C_o(u(k)) = w_w^H u(k). \]  
The Wiener weights are given by
\[ w_w = R^{-1} p, \]  
where \(R = E[u(k)u(k)^*] \) and \(p = E[u(k)d^*(k)]\), with MSE \( J_w = E[|d(k) - w_w^H u(k)|^2] \).

So under the independence and Gaussian assumptions, the LMS MSE is bounded by the Wiener MSE, i.e., \( J_{\text{LMS}} \geq J_w \). However, given that the LMS estimate in Eq. (3) is a function of much more information than the Wiener estimate in Eq. (6), this relation may not hold for realistic desired and reference signals, even allowing for the Gaussian assumption.

### 3. Optimum Linear Estimator

An absolute bound on the LMS estimator was presented in [6] and can be developed by assuming the processes \(\{d(k)\}\) and \(\{u(k)\}\) are jointly Gaussian, second-order, WSS, and observed over all time up to index \(k\).

Then the optimum MSE estimate
\[ \hat{d}_o(k) = C_o(u(k), u(k-1), \ldots; u(-\infty); d(k-1), \ldots; d(-\infty)), \]  
of \(d(k)\) is linear and can be written as
\[ \hat{d}_o(k) = \sum_{l=-\infty}^k h_u(k-l)u(l) + \sum_{l=-\infty}^{k-1} h_d(k-1-l)d(l), \]  
where \(h_u(l)\) and \(h_d(l)\) are causal filter impulse responses.

With a stability condition on \(h_u(l)\) and \(h_d(l)\), the optimum MSE of this linear estimator is given by [6]
\[ J_o = \int_{-1/2}^{1/2} S_d(f) df - \int_{-1/2}^{1/2} \left| H_d(f) \right|^2 S_d(f) df \]  
\[-2 Re \left\{ \int_{-1/2}^{1/2} e^{-2\pi f} H_d(f) H_u^*(f) S_u(f) df \right\} \]  
\[-\int_{-1/2}^{1/2} \left| H_u(f) \right|^2 S_u(f) df, \]  
where \(H_d(f)\) and \(H_u(f)\) are the frequency responses of \(h_d(l)\) and \(h_u(l)\). \(H_d(f)\) and \(H_u(f)\) are solutions of the integral equation
\[ \int_{-1/2}^{1/2} e^{2\pi fm} \left( \Gamma(f) - \Phi(f) H(f) \right) df \quad \forall m \geq 0, \]  
where
\[ \Phi(f) = \begin{bmatrix} S_u(f) & S_u(f) e^{-2\pi f} \\ S_u(f) e^{-2\pi f} & S_d(f) \end{bmatrix}, \]  
\[ \Gamma(f) = \begin{bmatrix} S_{du}(f) & S_d(f) e^{2\pi f} \\ S_d(f) e^{2\pi f} & S_d(f) \end{bmatrix}, \]  
\[ H(f) = \begin{bmatrix} H_u(f) & H_d(f) \end{bmatrix}^T, \]
and \(T\) denotes the matrix transpose. \(S_d(f)\) and \(S_u(f)\) are the power spectral densities (PSDs) of the processes \(\{d(k)\}\) and \(\{u(k)\}\) respectively, and \(S_{du}(f)\) is the cross spectrum between \(\{d(k)\}\) and \(\{u(k)\}\).

### 4. Transfer Function Approximation

An approach to modeling LMS adaptive filters that accounts for strong temporal correlation in the reference process was presented in [5] and was derived from the transfer function method originally proposed in [7, 8]. The technique is based upon constraining the processes \(\{d(k)\}\) and \(\{u(k)\}\) to be jointly WSS so that the Wiener filter is time invariant. By assuming that the LMS estimator approximately is in steady state at \(k = 0\) and
that $w(0) = w_w$ in Eq. (2), the error process can be written as [5]

$$e(k) + \sum_{j=0}^{k-1} u(j)e(j)u^H(j)u(k) = \varepsilon_w(k), \quad (13)$$

where $\varepsilon_w(k) = d(k) - w^H u(k)$ is the residual Wiener error process. This stochastic difference equation can be approximated by a standard difference equation by assuming that the reference process is also ergodic so that the second-order moments can be estimated from time averages. We then apply the estimate

$$u^H(j)u(k) \approx L r_u(k-j), \quad (14)$$

where $r_u(m)$ is the autocorrelation function of the reference process. Then Eq. (13) is approximated by

$$e(k) + \hat{\mu} L \sum_{j=0}^{k-1} r_u(k-j)e(j) = \varepsilon_w(k), \quad (15)$$

where $\hat{\mu} = \mu$ for LMS and $\hat{\mu} = \mu / L r_u(0)$ for NLMS. So the LMS error process can be interpreted as the output of a linear system with transfer function $H_E(z)$ given by [5, 8]

$$H_E(z) = \left\{ \begin{array}{ll}
1 & \text{LMS} \\
1 + \frac{\mu L R(z)}{1 + \mu R(z)/r_u(0)} & \text{NLMS},
\end{array} \right. \quad (16)$$

where

$$R(z) = \sum_{m=1}^{\infty} r_u(m)z^{-m}, \quad (17)$$

and driven by the WSS Wiener error process.

With this interpretation, the PSD of $\{e(k)\}$ is given by

$$S_e(f) = |H_E(f)|^2 S_{\varepsilon_w}(f), \quad (18)$$

where the PSD of $\{\varepsilon_w(k)\}$ is given by

$$S_{\varepsilon_w}(f) = S_d(f) - 2Re\{W_w(f)S_{\varepsilon_w}(f)\} + |W_w(f)|^2 S_u(f), \quad (19)$$

and $W_w(f)$ is the frequency response of the conjugate Wiener weights $w_w^*(l)$. Then the steady-state LMS MSE is the power of $\{e(k)\}$ given by

$$J_{\text{LMS}} = \int_{-1/2}^{1/2} S_e(f)df. \quad (20)$$

The PSD of the process $\{y(k)\}$ is approximated in the same way, and is given by

$$S_y(f) = S_d(f) - 2Re\{H_E(f)(S_d(f) - W_w(f)S_{\varepsilon_w}(f))\} + |H_E(f)|^2 S_{\varepsilon_w}(f). \quad (21)$$

5. Noise Canceler Problem

To demonstrate both the performance bound given in Sec. 3, and the approximation given in Sec. 4, consider the ANC configuration shown in Fig. 1. Let the reference signal $\{u(k)\}$, and desired signal $\{d(k)\}$, each be first-order auto regressive (AR) processes generated from the same white Gaussian noise source $\{v(k)\}$, with an independent Gaussian noise component added to each signal as shown in Fig. 2. The filter weights are updated with NLMS.

From [6], the reference and desired signals meet the conditions necessary to use Eq. (10) as a lower bound on the MSE. To model the NLMS ANC performance using the transfer function approach of Sec. 4, we calculate the Wiener filter of Eq. (7) using the auto and cross correlation functions of the reference and desired signals given by

$$r_d(m) = \frac{\sigma^2 u}{1 - r_d e^{-i\sigma}} \frac{r_{uv}|m|e^{i\sigma m}}{1 - r_u e^{-i\sigma}} \quad (22)$$

$$r_u(m) = \frac{\sigma^2 u}{1 - r_u e^{-i\sigma}} \frac{r_{uv}|m|e^{i\sigma m}}{1 - r_u e^{-i\sigma}} \quad (23)$$

and

$$r_{du}(m) = \left\{ \begin{array}{ll}
\frac{\sigma^2 u}{1 - r_{du} e^{-i\sigma}} \frac{r_{uv}|m|e^{i\sigma m}}{1 - r_u e^{-i\sigma}} & m > 0 \\
\frac{\sigma^2 u}{1 - r_{du} e^{-i\sigma}} \frac{r_{uv}|m|e^{i\sigma m}}{1 - r_u e^{-i\sigma}} & m \leq 0.
\end{array} \right. \quad (24)$$

Using the auto and cross spectral densities of the desired and reference processes

$$S_d(f) = \left| \frac{\sigma_u}{1 - r_d e^{-i\sigma}} e^{-i\sigma f} \right|^2 + \sigma_d^2, \quad (25)$$

$$S_u(f) = \left| \frac{\sigma_u}{1 - r_u e^{-i\sigma}} e^{-i\sigma f} \right|^2 + \sigma_u^2, \quad (26)$$

and

$$S_{du}(f) = \frac{\sigma^2 u}{(1 - r_{du} e^{-i\sigma})(1 - r_u e^{-i\sigma})}, \quad (27)$$

we form the PSD of the Wiener error process from Eq. (19). This is then used in Eq. (20) to obtain, through a numerical integration, the steady state MSE of the ANC.

Fig. 3 contains a plot of MSE as a function of the NLMS adaptation constant $\mu$. Included are curves representing performance of the optimal estimator from Eq. (10),
which forms a lower bound, and the transfer function approximation from Eq. (20). The AR poles have radii \( r = 0.99 \), with angle separation 3.6 degrees. The signal-to-noise ratios on the desired and reference channels are \( \text{SNR}_d = \text{SNR}_n = 20 \text{dB} \). The filter length is \( L = 25 \) taps. Notice first that the NLMS estimator does indeed outperform the finite Wiener filter. In addition, the performance of the transfer function model is close to and has the same form as that of the actual performance. Also, the NLMS estimator MSE performance comes close to achieving that of the optimal estimator as \( \mu \) increases to an optimal value. These curves demonstrate bounding and modeling techniques that provide insight into how the LMS estimator achieves its performance, and the limits on this performance.

Fig. 3 also displays MSE performance of the exponentially weighted RLS algorithm [1, pp. 566-571] versus the forgetting factor \( \lambda \). One possible explanation for this behavior is that the RLS algorithm is constrained to estimate the finite Wiener weights rather than directly minimize MSE as in the LMS algorithm, so its performance is tied more closely to that of the finite Wiener filter whose performance it does not exceed.

6. Equalizer Problem

Fig. 4 is a schematic of the AEQ structure. The AEQ is symmetric with \( N \) precursor and postcursor taps, where \( L = 2N + 1 \). The reference vector is comprised of three statistically independent components as

\[
\mathbf{u}(k) = \mathbf{s}(k) + \mathbf{x}(k) + \mathbf{n}(k),
\]

where \( \mathbf{s}(k) \) is the communication signal with power \( \sigma_s^2 \), \( \mathbf{x}(k) \) is the interference with power \( \sigma_x^2 \), and \( \mathbf{n}(k) \) is the noise with power \( \sigma_n^2 \). The reference sample \( \mathbf{u}(k) \) resides at the center tap at time \( k \). The communication signal and noise are modeled as zero-mean white processes. The signal symbols \( \mathbf{s}(k) \) are quadrature phase shift keyed (QPSK) with the mutually independent in-phase and quadrature components taking values \( +1 \) and \( -1 \) with equal probability (\( \sigma_s^2 = 2 \)). The noise and interference are Gaussian. The narrowband interference process has the ideal brickwall PSD

\[
S_z(f) = \begin{cases} 
N_z & f_\Delta - B_n/2 \leq f \leq f_\Delta + B_n/2 \\
0 & \text{otherwise}
\end{cases}
\]

(29)

where \( B_n \) is the normalized bandwidth and \( f_\Delta \) is the carrier offset frequency with \( \sigma_s^2 = N_z B_n \). The desired samples \( d(k) \) in Fig. 4 are equal to \( s(k) \) during training, and during decision-directed mode are equal to the output of the decision device.

Fig. 5 is a plot of MSE versus the NLMS step size parameter \( \mu \) for \( L = 51 \), signal-to-noise ratio 25dB (SNR = \( \sigma_s^2/\sigma_n^2 \)), and signal-to-interference ratio -10dB (SIR = \( \sigma_s^2/\sigma_x^2 \)). The interference is quite narrow with normalized bandwidth \( B_n = 0.02 \) and offset frequency \( f_\Delta = 0.25 \). Theoretical NLMS MSE derived numerically from Eq. (20) is compared to experimental MSE determined through repeated simulations during the steady-state, decision-directed mode. The interference is simulated using a high-order Chebyshev filter. Also plotted is the MSE of the finite Wiener filter and a decision feedback equalizer (DFE), which is identical in structure to the Wiener filter except that it also filters previously detected symbols using a 51-tap feedback filter. In principle, the DFE is equivalent to the optimum linear estimator of the ANC of Sec. 5, except that the DFE has finite length, and \( s(k) \) is not Gaussian. Finally, the experimental MSE for exponentially weighted RLS versus \( \lambda \) is plotted. For stability reasons, RLS operates in the training mode.

Clearly, with the proper choice of \( \mu \), NLMS can outperform the finite Wiener filter for this scenario, and in fact approaches the MSE performance of the DFE without explicitly incorporating a feedback filter. This figure is interesting because it contradicts conventional adaptive filter theory in which small \( \mu \) is associated with small MSE. It also is interesting that, as in the ANC example, this behavior is not observed with RLS, even though the same information is used in both the RLS and NLMS estimators. However, it is possible that an extended RLS algorithm can be devised that more effectively incorporates the feedback information for the ANC and AEQ problems [9].

To demonstrate that this effect diminishes as the bandwidth of the interferer increases, Fig. 6 is a plot of MSE versus \( B_n \). The NLMS step size is fixed at \( \mu = 0.60 \) for all bandwidths and \( f_\Delta = 0 \). Closer agreement with theory is observed at the narrow bandwidths than the broader ones. Also, as \( B_n \) approaches the bandwidth of the additive noise \( B_n - 1 \), the ‘non-Wiener’ effect disappears and we observe conventional ‘excess’ MSE.

Fig. 7 is a sequence of estimated PSDs of the NLMS output process \( \{ y(k) \} \) for the scenario corresponding to Fig. 5 with three values of \( \mu \). Also plotted are the theoretical PSDs obtained using Eq. (21). Fig. 7(a) represents ‘small’ \( \mu \) for which NLMS MSE is similar to Wiener MSE in Fig. 5. There is a notching effect that nulls the interference at the expense of distorting the desired signal \( \{ s(k) \} \). In contrast, Fig. 7(b) is the result of using \( \mu = 0.60 \), which corresponds to the minimum of the MSE curve in Fig. 5. The PSD is seen to be flat, which suggests that interference nulling is occurring without distorting \( \{ s(k) \} \). Fig. 7(c) is for ‘large’ \( \mu \) where spectral distortion is observed again.
7. Conclusions

We demonstrated that LMS adaptive filters that operate in environments consisting of mixtures of broadband and narrowband processes can outperform the corresponding Wiener filter in terms of MSE. This effect occurs because the LMS estimate is a function of more information than used by the Wiener filter. Although all of the processes considered are WSS, this phenomenon is reminiscent of the tracking problem involving non-stationary processes. However, the MSE of the LMS estimator is bounded by a time-invariant linear Wiener filter that explicitly uses the same information. Also, RLS does not exhibit this behavior for the considered interference canceling problems, emphasizing the importance of the form of the estimate rather than just the information.

8. References


Figure 1: Adaptive noise canceler.

Figure 2: Autoregressive interference model.

Figure 3: ANC MSE of NLMS as a function of μ and RLS as a function of λ with SNRd = 20dB, SNRn = 20dB, τ = 0.99, α − β = 3.6deg, and L = 25.
Figure 4: Adaptive equalizer structure.

Figure 5: Equalizer MSE of NLMS as a function of $\mu$ and RLS as a function of $\lambda$ for an interference with bandwidth $B_n = 0.02$ and offset frequency $f_\Delta = 0.25$, with $L = 51$, SNR = 25dB, and SIR = -10dB.

Figure 6: Equalizer MSE of NLMS as a function of interferer bandwidth $B_n$ with offset frequency $f_\Delta = 0$, using $\mu = 0.60$, with $L = 51$, SNR = 25dB, and SIR = -10dB.

Figure 7: Estimated and theoretical power spectral densities of NLMS equalizer output process $\{y(k)\}$ for (a) $\mu = 0.04$, (b) $\mu = 0.60$, (c) $\mu = 1.92$, and remaining parameters as in Fig. 5.