Stochastic Analysis of the Filtered-X LMS Algorithm in Systems With Nonlinear Secondary Paths

Márcio H. Costa, José Carlos M. Bermudez, *Member, IEEE,* and Neil J. Bershad, *Fellow, IEEE*

***Abstract*—This paper presents a statistical analysis of the filtered-X LMS algorithm behavior when the secondary path (output of the adaptive filter) includes a nonlinear element. This system is of special interest for active acoustic noise and vibration control, where a saturation nonlinearity models the nonlinear distortion introduced by the power amplifiers and transducers. Deterministic nonlinear recursions are derived for Gaussian inputs for the transient mean weight, mean square error, and cross-covariance matrix of the adaptive weight vector at different times. The cross-covariance results provide improved steady-state predictions (as compared with previous results) for moderate to large step sizes. Monte Carlo simulations show excellent agreement with the behavior predicted by the theoretical models. The analytical and simulation results show that a small nonlinearity can have a significant impact on the adaptive filter behavior.**

***Index Terms*—Adaptive filters, adaptive signal processing, least mean square methods, transient analysis.**

I. Introduction

ADAPTIVE algorithms are applicable to system identification and modeling, noise and interference cancelling, equalization, signal detection, and prediction [1], [2]. Most adaptive system analyses neglect nonlinear effects and model the unknown systems as linear with memory. In many important practical circumstances, a linear model simplifies the mathematics and permits detailed system analysis. More sophisticated models must be used when nonlinear effects significantly impact actual system behavior [3]–[5]. Important nonlinear effects occur in active noise control (ANC) and active vibration control (AVC) systems, for example, [4]. ANC and AVC systems include acoustical/mechanical paths. Signal converters (A/D and D/A), power amplifiers, and transducers (speakers or actuators) can nonlinearly transform digital electrical signals into analog electrical or mechanical signals. This nonlinear effect is caused by overdriving the electronics or the speakers/transducers in the secondary path[1] [6], [7]. Thus, the nonlinearities should be included in the mathematical model for accurate analysis.

Bernhard *et al.* [8] briefly discussed such nonlinear effects but presented no analysis. Costa *et al.* [9], [10] recently studied the statistical behavior of the LMS algorithm for a memoryless nonlinear secondary path. Small nonlinearities were shown to significantly affect algorithm performance. These analytical results provide important insights of the effect of a nonlinear secondary path upon ANC and AVC system behavior. However, they do not provide information about secondary path impulse response effects.

This paper provides a stochastic analysis of the FXLMS algorithm with a saturation nonlinearity as shown in Fig. 1. The function $g(y_s)$ is a zero-memory saturation nonlinearity.[2] S is the secondary path linear filter and $\hat{S}$ is its estimate. Usually, $\hat{S}$ is designed to duplicate S. The mismatched case is analyzed here since perfect estimation cannot be achieved in practice. A degree of nonlinearity is defined that measures the impact of the nonlinearity on the achievable mean square error (MSE). Deterministic nonlinear recursions are derived for Gaussian inputs and slow adaptation for the transient mean weight, mean square error, and cross-covariance matrix of the adaptive weight vector at different times. The analytical and simulation results show that even a small nonlinearity can have a significant impact on the adaptive filter behavior. Contrary to the case studied in [10], the converged mean weight vector is not a scaled version of the primary path response. The steady-state solution depends on the nonlinearity $g(\cdot)$, the secondary path impulse response S, and the secondary path impulse response estimate $\hat{S}$. The cross-covariance results provide improved steady-state predictions (as compared with previous results) for moderate to large step sizes. The models presented here generalize the linear case analyses in [11]–[13]. The models and the results for the linear case correspond to a degree of nonlinearity equal to zero. A wide variety of Monte Carlo simulations show excellent agreement with the theoretical predictions.

II. Analysis—FXLMS Algorithm Transient Behavior

A. Problem Definition

Fig. 1 shows a block diagram for the FXLMS algorithm [1] with a nonlinearity at the output of the adaptive filter. The

Manuscript received August 13, 2001; revised February 26, 2002. This work was supported in part by the Brazilian Ministry of Education (CAPES) under Grant PICDT 0129/97-9 and by the Brazilian Ministry of Science and Technology (CNPq) under Grant 352084/92-8. The associate editor coordinating the review of this paper and approving it for publication was Dr. Naofal Al-Dhahir.

M. H. Costa is with the Grupo de Engenharia Biomédica, Escola de Engenharia e Arquitetura, Universidade Católica de Pelotas, Pelotas, Brazil (e-mail: costa@atlas.ucpel.tche.br).

J. C. M. Bermudez is with the Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis, Brazil (e-mail: bermudez@fastlane.com.br).

N. J. Bershad is with the Department of Electrical and Computer Engineering, University of California Irvine, Irvine, CA 96032 USA (e-mail: bershad@ece.uci.edu).

Publisher Item Identifier S 1053-587X(02)04383-0.

[1]Secondary path is the usual term for the path leading from the adaptive filter output to the cancellation point [1].

[2]The modeling of nonlinear effects in amplifiers and tranducers is very complex, and there is no unique model for all situations [6]–[8]. Static nonlinearities have been used to model nonlinear effects in electronics and in transducers [3], [5].

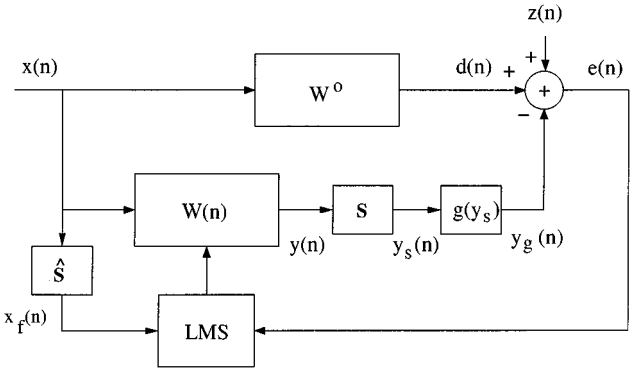

Fig. 1. Block diagram of the nonlinear adaptive system.

notation used is as follows: $W^o = [w_0^o\, w_1^o \cdots w_{N-1}^o]^T$ is the response of the unknown system; $W(n) = [w_0(n)\, w_1(n) \cdots w_{N-1}(n)]^T$ is the adaptive filter weight vector; $S = [s_0\, s_1 \cdots s_{M-1}]^T$ is the secondary path response; $\hat{S} = [\hat{s}_0\, \hat{s}_1 \cdots \hat{s}_{\hat{M}-1}]^T$ is the estimate of the secondary path response; $x(n)$ is the reference signal; $x_f(n)$ is the filtered reference signal; $X(n) = [x(n)\, x(n-1) \cdots x(n-N+1)]^T = [x_0(n)\, x_1(n) \cdots x_{N-1}(n)]^T$ is the observed input data vector; $X_f(n) = [x_f(n)\, x_f(n-1) \cdots x_f(n-N+1)]^T = [x_{f_0}(n)\, x_{f_1}(n) \cdots x_{f_{N-1}}(n)]^T$ is the filtered input data vector; $d(n)$ is the primary signal; $z(n)$ is the measurement noise; $y(n)$ is the output of the adaptive filter; $y_s(n)$ is the output of the linear filter S; $g(y_s)$ is the saturation nonlinearity; $y_g(n)$ is the nonlinearity output; and $e(n)$ is the error signal.

In the following analysis, $x(n)$ is assumed stationary, zero-mean, and Gaussian with variance σ_x^2. The measurement noise $z(n)$ is stationary, white, zero-mean, Gaussian, with variance σ_z^2 and statistically independent of any other signal. The saturation nonlinearity is modeled by the scaled error function

$$g(y) = \int_0^y e^{-(z^2/2\sigma^2)}\, dz. \tag{1}$$

Note that $\lim_{\sigma^2\to\infty}[g(y)] = y$ and $\lim_{\sigma^2\to 0}[g(y)] = \sigma\sqrt{\pi/2}\ \mathrm{sgn}(y)$. Hence, by changing σ^2, $g(y)$ can be varied between a linear device and a hard limiter. The effects of very large nonlinearities ($\sigma \to 0$) can be studied by scaling $g(y)$ by a constant such as A/σ, $A \in \mathbb{R}^+$. $g(y)$ models the saturation type nonlinearity, which is of great practical interest.

The degree of nonlinearity η^2 for the system in Fig. 1 is defined as

$$\eta^2 = \frac{\pi}{2}\,\frac{E\,\{y_s^2(n)\}\big|_{\substack{n\to\infty\\ \sigma^2\to\infty}}}{\max\{y_g^2(n)\}} = \frac{1}{\sigma^2}\,E\,\{y_s^2(n)\}\big|_{\substack{n\to\infty\\ \sigma^2\to\infty}}. \tag{2}$$

η^2 is a steady-state parameter that describes the power limitation in $y_g(n)$ due to the nonlinear distortion. The denominator of (2) is the maximum nonlinear output power and is $\max\{y_g^2(n)\} = (\pi/2)\sigma^2 = \lim_{y\to\infty}[g(y)]$. The numerator of (2) is the average steady-state power of the cancelling signal in the linear case.[3]

[3]In the case, (1) is scaled by A/σ, $\max\{y_g^2(n)\} = (\pi/2)A^2$, and (2) becomes

$$\eta^2 = \frac{\pi}{2}\,\frac{E\,\{y_s^2(n)\}\big|_{\substack{n\to\infty\\ \sigma^2\to\infty}}}{\max\{y_g^2(n)\}} = \frac{1}{A^2}\,E\,\{y_s^2(n)\}\big|_{\substack{n\to\infty\\ \sigma^2\to\infty}}.$$

The adaptive weight update equation for the FXLMS algorithm is [1]

$$W(n+1) = W(n) + \mu e(n) X_f(n) \tag{3}$$

where μ is the adaptation step size. The filtered input data vector $X_f(n)$ can be written as

$$X_f(n) = \sum_{j=0}^{\hat{M}-1} \hat{s}_j X(n-j). \tag{4}$$

The error signal is given by

$$\begin{aligned} e(n) &= d(n) + z(n) - y_g(n) \\ &= W^{o^T} X(n) + z(n) \\ &\quad - g\left[\sum_{i=0}^{M-1} s_i W^T(n-i) X(n-i)\right]. \end{aligned} \tag{5}$$

Substituting (4) and (5) in (3) yields

$$\begin{aligned} W(n+1) = W(n) + \mu \Bigg\{ & W^{o^T} X(n) + z(n) \\ & - g\left[\sum_{i=0}^{M-1} s_i W^T(n-i) X(n-i)\right] \Bigg\} X_f(n). \end{aligned} \tag{6}$$

B. Statistical Assumptions

Equations (4)–(6) clearly show that the statistical analysis of the algorithm behavior involves moments of products of the present and past values of data and weight vectors. Since the joint probability density function of the weights and data is not known, some statistical approximations must be made to proceed with the analysis. The following assumptions are used for sufficiently small μ:

A1)

$$\begin{aligned} &E\{W^T(n-i) X(n-i) X^T(n-j) W(n-j)\} \\ &\qquad \approx E\{W^T(n-i) R_{j-i} W(n-j)\} \end{aligned}$$

A2)

$$\begin{aligned} &E\{X_f(n) X^T(n) | \Psi\} \\ &\qquad = E\left\{\sum_{j=0}^{\hat{M}-1} \hat{s}_j X(n-j) X^T(n) | \Psi\right\} \approx \sum_{j=0}^{\hat{M}-1} \hat{s}_j R_{-j} \end{aligned}$$

where $\Psi = \{W(n),\ W(n-1), \ldots, W(n-M+1)\}$, and $R_{k-\ell} = E\{X(n-\ell) X^T(n-k)\}$ is the correlation matrix of time-lagged input vectors.

A sufficient condition for these assumptions to hold is that weight and data vectors are statistically independent. Clearly, this is not true because of the different time indices on W and X [i.e., future values of W, say $W(q)$, depend on past values of X, say, $X(r)$, for $r < q$]. A1 and A2 imply only that the statistical dependence of weight and data vectors is not as significant in determining the algorithm behavior as the dependence between lagged input vectors. Assumptions A1 and A2 are supported by extensive numerical simulations and by previous results [10], [11]. Some sample simulations are given in the examples at the end of the paper.

C. Mean Weight Behavior

The expected value of (6) will be evaluated in two steps. First, the expectation is taken conditioned on Ψ, leading to

$$\begin{aligned}E\{W(n+1)|\Psi\} &= W(n)+\mu E\left\{X_f(n)X^T(n)|\Psi\right\}W^o \\ &\quad +\mu E\left\{z(n)X_f(n)|\Psi\right\} \\ &\quad -\mu E\left\{g\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]X_f(n)|\Psi\right\}.\end{aligned} \tag{7}$$

The first expectation in (7) is evaluated using A2. The second expectation in (7) is zero since $z(n)$ is zero-mean and statistically independent of $X(n)$ [and, consequently, of $X_f(n)$]. The third expectation is of the form $E\{g(y_1)Y_2\}$, where y_1 and the components of vector Y_2 are zero-mean Gaussian variates. This expectation can be determined using the results in [14, Eq. A19] with $b_1 = 0$, $c = 1/\sigma$, $\sigma_q = \sigma$, $W^TX = \sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)$, $X = X_f(n)$, $R_X W = \sum_{i=0}^{M-1}\sum_{j=0}^{\hat{M}-1} s_i\hat{s}_j R_{i-j}W(n-i)$, and $\sigma_y^2 = \sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)$. Thus

$$\begin{aligned}E\{W(n+1)|\Psi\} &= W(n)+\mu\sum_{j=0}^{\hat{M}-1}\hat{s}_j R_{-j}W^o \\ &\quad -\mu\frac{1}{\sqrt{\frac{1}{\sigma^2}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)+1}} \\ &\quad \cdot\sum_{i=0}^{M-1}\sum_{j=0}^{\hat{M}-1} s_i\hat{s}_j R_{i-j}W(n-i).\end{aligned} \tag{8}$$

The expected value of (8) can only be approximated since the joint probability density function of $W(n)$ and $W(m)$ is not known. A good approximation is obtained by noticing that $\sum_{k=0}^{N-1} w_k(n-j)w_k(n-i)$ can be assumed weakly correlated to $w_\ell(n-i)$ for large values of N and for all i, j, and ℓ. This is equivalent to applying the averaging principle proposed in [15] as the value of the summation tends to be slowly varying when compared with $w_\ell(n-i)$ for N large. Approximating $W(n-i)$ and $W^T(n-j)R_{i-j}W(n-i)$ by their expected values separately in the numerator and denominator of (8)[4] and using

$$\begin{aligned}E\left\{W^T(n-j)R_{i-j}W(n-i)\right\} \\ = \mathrm{tr}\left[R_{i-j}E\left\{W(n-i)W^T(n-j)\right\}\right]\end{aligned} \tag{9}$$

where tr[·] is the trace of the matrix

$$\begin{aligned}E\{W(n+1)\} \approx E\{W(n)\} &+\mu\sum_{j=0}^{\hat{M}-1}\hat{s}_j R_{-j}W^o \\ &-\mu\frac{1}{\sqrt{\frac{1}{\sigma^2}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i \mathrm{tr}[R_{i-j}K_{i,j}(n)]+1}} \\ &\cdot\sum_{i=0}^{M-1}\sum_{j=0}^{\hat{M}-1} s_i\hat{s}_j R_{i-j}E\{W(n-i)\}\end{aligned} \tag{10}$$

[4]This approximation has also been successfully applied in [10].

where

$$\begin{aligned}K_{i,j}(n) &= E\{W(n-i)W^T(n-j)\} \\ &\quad i\geq 0 \quad \text{and} \quad j\geq 0.\end{aligned} \tag{11}$$

Note that as $\sigma^2\to\infty$, (10) reduces to the mean weight equation derived in [11, Eq. 16] for the linear case. In a previous publication [16], $K_{i,j}(n) \approx E\{W(n-i)\}E\{W^T(n-j)\}$ was assumed. This simplified model accurately predicts the algorithm behavior for small μ. For larger μ, the simplified model is not accurate in steady state. Hence, an approximate recursive expression for $K_{i,j}(n)$ will be derived in Appendix A.

D. Mean Square Error (MSE) Behavior

Squaring (5) and taking the conditional expectation given Ψ yields

$$\begin{aligned}E\{c^2(n)|\Psi\} &= W^{o^T}E\{X(n)X^T(n)|\Psi\}W^o + 2W^{o^T}E\{z(n)X(n)|\Psi\} \\ &\quad -2W^{o^T}E\left\{g\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]X(n)\middle|\Psi\right\} \\ &\quad +E\{z^2(n)|\Psi\} \\ &\quad -2E\left\{z(n)g\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]\middle|\Psi\right\} \\ &\quad +E\left\{g^2\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]\middle|\Psi\right\}.\end{aligned} \tag{12}$$

Using A2, the first expectation equals R_0. The second and fifth expectations are zero since $z(n)$ is i.i.d. and zero mean. The fourth expectation is given by $E\{z^2(n)\} = \sigma_z^2$.

Following the same steps used to derive the third term in the right-hand part of (10) (with $\hat{s}_0 = 1$ and $\hat{s}_j = 0$, $j\neq 0$), the third expectation of (12) is

$$\begin{aligned}E\left\{g\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]X(n)|\Psi\right\} \\ = \frac{1}{\sqrt{\frac{1}{\sigma^2}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)+1}} \\ \cdot\sum_{i=0}^{M-1} s_i R_i W(n-i).\end{aligned} \tag{13}$$

The last expectation in (12) has the form $E\{g^2(y)\}$, where y is zero-mean Gaussian. Using the results in [17, Eq. 40], with $\alpha_f = b_1(n) = 1$, $H^TY = \sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)$, and $b = \sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)$, yields

$$\begin{aligned}E\left\{g^2\left[\sum_{i=0}^{M-1} s_i W^T(n-i)X(n-i)\right]|\Psi\right\} \\ = \sigma^2\sin^{-1}\left(\frac{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)}{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j)R_{i-j}W(n-i)+\sigma^2}\right).\end{aligned} \tag{14}$$

Using the above results in (12) yields

$$
\begin{aligned}
&E\{e^2(n)|\Psi\} \\
&\quad = W^{o^T} R_0 W^o + \sigma_z^2 \\
&\qquad - \frac{2}{\sqrt{\frac{1}{\sigma^2}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j) R_{i-j} W(n-i) + 1}} \\
&\qquad \cdot W^{o^T} \sum_{i=0}^{M-1} s_i R_i W(n-i) + \sigma^2 \sin^{-1} \\
&\qquad \cdot \left(\frac{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j) R_{i-j} W(n-i)}{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i W^T(n-j) R_{i-j} W(n-i) + \sigma^2} \right).
\end{aligned} \tag{15}
$$

Again, approximating $W(n)$ and $W^T(n-j)R_{i-j}W(n-i)$ by their expected values, (15) leads to an expression for the MSE $\xi(n)$

$$
\begin{aligned}
\xi(n) &= E\{e^2(n)\} \\
&= W^{o^T} R_0 W^o + \sigma_z^2 \\
&\quad - \frac{2}{\sqrt{\frac{1}{\sigma^2}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i \mathrm{tr}[R_{i-j} K_{i,j}(n)] + 1}} \\
&\quad \cdot W^{o^T} \sum_{i=0}^{M-1} s_i R_i E\{W(n-i)\} + \sigma^2 \sin^{-1} \\
&\quad \cdot \left(\frac{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i \mathrm{tr}[R_{i-j} K_{i,j}(n)]}{\sum_{j=0}^{M-1}\sum_{i=0}^{M-1} s_j s_i \mathrm{tr}[R_{i-j} K_{i,j}(n)] + \sigma^2} \right).
\end{aligned} \tag{16}
$$

Equation (16) requires $K_{i,j}(n)$, as does (10). The MSE for the linear case [12, Eq. (11)] can be obtained by letting $\sigma^2 \to \infty$ in (16).

E. Weight Correlation Matrix

An approximate recursion for $E\{W(n-i)W^T(n-j)|\Psi\}$ is derived in Appendix A [Eq. (42)]. The expectation of (42) requires approximations because neither the joint density function of the weight vector at two different times nor the contributions of higher moments are known. An approximate expression for the expected value of (42) as a function of first and second moments of the weight vector components is given by

$$
\begin{aligned}
&K_{i,j}(n+1) \\
&\quad = K_{i,j}(n) + \mu E\{W(n-i)\} W^{o^T} \tilde{R}_{\hat{s}}^T \\
&\qquad + \mu \tilde{R}_{\hat{s}} W^o E\{W^T(n-j)\} \\
&\qquad - \frac{\mu \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1} s_p \hat{s}_q R_{p-q} K_{i+p,j}(n)}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p} K_{q,p}(n-i)] + 1}} \\
&\qquad - \frac{\mu \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1} s_p \hat{s}_q K_{i,j+p}(n) R_{p-q}^T}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p} K_{q,p}(n-j)] + 1}} \\
&\qquad + \mu^2 \sigma_z^2 \delta(i-j) \sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1} \hat{s}_p \hat{s}_q R_{j-i+q-p} \\
&\qquad + \mu^2 \sum_{p=0}^{\hat{M}-1} \hat{s}_p R_{-p} W^o W^{o^T} \sum_{q=0}^{\hat{M}-1} \hat{s}_q R_q \\
&\qquad + \mu^2 \sum_{p=0}^{\hat{M}-1} \hat{s}_p R_{j-i-p} W^o W^{o^T} \sum_{q=0}^{\hat{M}-1} \hat{s}_q R_{j-i+q} \\
&\qquad + \mu^2 W^{o^T} R_{j-i} W^o \sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1} \hat{s}_p \hat{s}_q R_{j-i+q-p} \\
&\qquad + \frac{\mu^2}{\sigma^2 \left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p} K_{q,p}(n-i)] + 1 \right)^{3/2}} \\
&\qquad \cdot \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1} s_r s_k \hat{s}_l s_p \hat{s}_q E\{W^T(n-i-r)\} \\
&\qquad \cdot R_{j-i-r} W^o R_{k-l} K_{k,p}(n-i) R_{j-i+q-p} \\
&\qquad + \frac{\mu^2}{\sigma^2 \left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p} K_{q,p}(n-j)] + 1 \right)^{3/2}} \\
&\qquad \cdot \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1} s_r s_k \hat{s}_l s_p \hat{s}_q E\{W^T(n-j-r)\} \\
&\qquad \cdot R_{i-j-r} W^o R_{i-j+q-p}^T K_{p,k}(n-j) R_{k-l}^T \\
&\qquad - \frac{\mu^2}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p} K_{q,p}(n-i)] + 1}} \\
&\qquad \cdot \Bigg[\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1} \hat{s}_p \hat{s}_q s_r E\{W^T(n-i-r)\} \\
&\qquad \cdot R_{j-i-r} W^o R_{j-i+q-p} \\
&\qquad + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1} s_p \hat{s}_q R_{p-q} E\{W(n-i-p)\} W^{o^T} \tilde{R}_{\hat{s}}^T \\
&\qquad + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{\hat{M}-1} s_p \hat{s}_q \hat{s}_r R_{j-i-r} W^o \\
&\qquad \cdot E\{W^T(n-i-p)\} R_{j-i+q-p} \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{\mu^2}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p}K_{q,p}(n-j)]+1}} \\
&\cdot \left[\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1} \hat{s}_p\hat{s}_q s_r E\{W^T(n-j-r)\}\right. \\
&\cdot R_{i-j-r}W^o R^T_{i-j+q-p} \\
&+ \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1} s_p\hat{s}_q \tilde{R}_{\hat{s}} W^o E\{W^T(n-j-p)\}R^T_{p-q} \\
&+ \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{\hat{M}-1} s_p\hat{s}_q\hat{s}_r R^T_{i-j+q-p} \\
&\left.\cdot E\{W(n-j-p)\}W^{o^T}R^T_{i-j-r}\right] \\
&+ \mu^2 E\left\{Q\left(S, \hat{S}, R_{k-\ell}, n, i, j\right)\right\} \qquad (17)
\end{aligned}
$$

where $\tilde{R}_{\hat{s}}$ was defined after (42). The expected values of terms (A) and (B) in (42) include third-order moments of the weights. Each of these moments were approximated by a product of first- and second-order moments. This approximation preserves the mean and fluctuation behaviors of $W(n)$ while keeping the mathematical problem tractable [10, Eq. (29)].

An expression for $E\{Q(S, \hat{S}, R_{k-\ell}, n, i, j)\}$ can be obtained from the evaluation of $E\{g(y_1)g(y_2)Y_3Y_4^T\}$, where y_1 and y_2 and the components of Y_3 and Y_4 are zero-mean jointly Gaussian variates. After laborious calculations, such an expression follows using the same approach as in [17, Eq. A6]. However, only a special case is needed to determine the algorithm behavior, as will be shown in Section II-F. Thus, the complete expression is not presented here to conserve space.

F. Recursive Generation of the Weight Correlation Matrix

The time delays of the weight vectors in (17) are functions of several parameters (i, j, p, q, l and r). Thus, it is not clear how to determine $K_{i,j}(n+1)$ recursively from previous available results. The following properties of the matrices $K_{a,b}(n)$ for any a, b, and $\Delta \in \mathbb{Z}$ are instrumental:

$$
\begin{cases}
K_{a,b}(n) = E\{W(n-a)W^T(n-b)\} \\
\quad = \left[E\{W(n-b)W^T(n-a)\}\right]^T = K^T_{b,a}(n) \\
K_{a-\Delta,b-\Delta}(n-\Delta) = E\{W(n-\Delta-a+\Delta) \\
\quad W^T(n-\Delta-b+\Delta)\} = K_{a,b}(n).
\end{cases} \qquad (18)
$$

Using (18), it is possible to write

$$
K_{a,b}(n) = \begin{cases} K_{0,b-a}(n-a), & b > a \\ K_{0,0}(n-a), & b = a \\ K_{a-b,0}(n-b) = K^T_{0,a-b}(n-b), & b < a. \end{cases} \qquad (19)
$$

Thus, $K_{a,b}(n)$ can be obtained for any $a \geq 0$ and $b \geq 0$ from present and previous values of $K_{0,k}$, $0 \leq k \leq \max\{M, \hat{M}\}$. Recursive expressions are now derived for $K_{0,0}(n)$ and $K_{0,k}(n)$, $k \neq 0$.[5]

For $i = j = 0$, the term $E\{Q(S, \hat{S}, R_{k-\ell}, n, i, j)\}$ in (17) becomes [17]

$$
E\left\{Q_{0,0}\left(S, \hat{S}, R_{k-\ell}, n\right)\right\} = E\left\{E\left\{g^2(y_1)Y_2Y_2^T|\Psi\right\}\right\} \qquad (20)
$$

where $y_1 = \sum_{k=0}^{M-1} s_k W^T(n-k)X(n-k)$, and $Y_2 = X_f(n)$. Using the results in [17, Eq. A13], it can be seen in (21), shown on the bottom of the page, where $\tilde{R}_{\hat{s}\hat{s}} = \sum_{i=0}^{\hat{M}-1}\sum_{j=0}^{\hat{M}-1} \hat{s}_i\hat{s}_j R_{i-j}$.

Setting $i = j = 0$ in (17) and using (19) and (21) yields the recursion shown in (22), shown at the bottom of the next page, for $K_{0,0}(n)$.

The final step is the derivation of a recursion for $K_{0,k}(n)$, $k \neq 0$. Post-multiplying a delayed version of (6) by $W^T(n-k-j+1)$ yields

$$
\begin{aligned}
&W(n-i-1)W^T(n-k-j+1) \\
&\quad = W(n-i)W^T(n-k-j+1) \\
&\qquad + \mu\left\{W^{o^T}X(n-i) + z(n-i)\right. \\
&\qquad\qquad \left. - g\left[\sum_{k=0}^{M-1} s_k W^T(n-i-k)X(n-i-k)\right]\right\} \\
&\qquad \cdot X_f(n-i)W^T(n-k-j+1). \qquad (23)
\end{aligned}
$$

[5] $K_{0,0}(n)$ cannot be obtained by making $k = 0$ in the recursive expression for $K_{0,k}(n)$. This would require $K_{a,b}(n)$ for $b < 0$. Thus, $K_{0,0}(n)$ must be obtained directly from (17).

$$
\begin{aligned}
&E\left\{Q_{0,0}\left(S, \hat{S}, R_{k-\ell}, n\right)\right\} \\
&\quad = \sigma^2 \sin^{-1}\left(\frac{\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p}K_{q,p}(n)]}{\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p}K_{q,p}(n)] + \sigma^2}\right)\tilde{R}_{\hat{s}\hat{s}} \\
&\qquad + \frac{2\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1}\sum_{k=0}^{\hat{M}-1} s_p\hat{s}_q s_r\hat{s}_k R_{p-q}K_{p,r}(n)R_{k-r}}{\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p}K_{q,p}(n)]+1\right)\sqrt{\frac{2}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q \mathrm{tr}[R_{q-p}K_{q,p}(n)]+1}} \qquad (21)
\end{aligned}
$$

Taking the expected value of (23) with the same assumptions as before and using the notation for $K_{a,b}(n)$ yields

$$K_{i,k+j}(n+1) = K_{i,k+j-1}(n) + \mu\tilde{R}_{\hat{s}}W^{o}E\{W^{T}(n-k-j+1)\} - \mu\frac{\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}K_{i+p,k+j-1}(n)}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{p-q}K_{p,q}(n-i)]+1}} \quad (24)$$

which yields for $i=j=0$

$$K_{0,k}(n+1) = K_{0,k-1}(n) + \mu\tilde{R}_{\hat{s}}W^{o}E\{W^{T}(n-k+1)\} - \mu\frac{\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}K_{p,k-1}(n)}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{p-q}K_{p,q}(n)]+1}}. \quad (25)$$

For $k=1,\ldots,M-1$, (22) and (25) constitute a set of M recursions that can be used to determine the weight vector correlation matrices $K_{i,j}(n)$ required for the recursive evaluation of mean weight behavior [see (10)] and the MSE behavior [see (16)].

The complete analytical model is composed of (10), (16), (19), (22), and (25). Appendix B presents a step-by-step procedure to initialize the variables and to recursively evaluate $E\{W(n)\}$ and $\xi(n)$ using the model. Since its implementation is quite laborious, a Matlab code is also available at the authors' web site for download.[6]

III. Steady-State Algorithm Behavior

This section studies the limiting behavior of the converged FXLMS algorithm. The determination of the steady-state algorithm behavior from (10), (16), (22), and (25) requires numerical methods. However, using the assumption of very small weight fluctuations (compared to their mean values), simple analytical expressions that are useful for evaluation and design purposes can be determined. Thus, $W(n)\approx E\{W(n)\}$ is assumed for the steady-state analysis.

[6]http://www.eel.ufsc.br/~bermudez. The code has not been optimized for speed but has been tested in numerous examples for reliability and correspondence to the analytical model.

$$\begin{aligned}
K_{0,0}(n+1) = {} & K_{0,0}(n) + \mu E\{W(n)\}W^{o^T}\tilde{R}_{\hat{s}}^{T} + \mu\tilde{R}_{\hat{s}}W^{o}E\{W^{T}(n)\} + \mu^2\sigma_z^2\tilde{R}_{\hat{s}\hat{s}} + 2\mu^2\tilde{R}_{\hat{s}}W^{o}W^{o^T}\tilde{R}_{\hat{s}}^{T} \\
& + \mu^2W^{o^T}R_0W^{o}\tilde{R}_{\hat{s}\hat{s}} - \frac{\mu\left[\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}K_{0,p}^{T}(n) + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qK_{0,p}(n)R_{p-q}^{T}\right]}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+1}} \\
& + \frac{2\mu^2\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1}s_p\hat{s}_qs_k\hat{s}_ls_rE\{W^{T}(n-r)\}R_{-r}W^{o}R_{k-l}K_{p,k}^{T}(n)R_{q-p}}{\sigma^2\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+1\right)^{3/2}} \\
& - \frac{2\mu^2}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+1}}\cdot\left\{\sum_{r=0}^{M-1}s_rE\{W^{T}(n-r)\}R_{-r}W^{o}\tilde{R}_{\hat{s}\hat{s}}\right. \\
& \left. + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}E\{W(n-p)\}W^{o^T}\tilde{R}_{\hat{s}}^{T} + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_q\tilde{R}_{\hat{s}}W^{o}E\{W^{T}(n-p)\}R_{q-p}\right\} \\
& + \mu^2\sigma^2\sin^{-1}\left(\frac{\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]}{\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+\sigma^2}\right)\tilde{R}_{\hat{s}\hat{s}} \\
& + \frac{2\mu^2\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1}\sum_{k=0}^{\hat{M}-1}s_p\hat{s}_qs_r\hat{s}_kR_{p-q}K_{p,r}(n)R_{k-r}}{\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+1\right)\sqrt{\frac{2}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_q\mathrm{tr}[R_{q-p}K_{q,p}(n)]+1}}
\end{aligned} \quad (22)$$

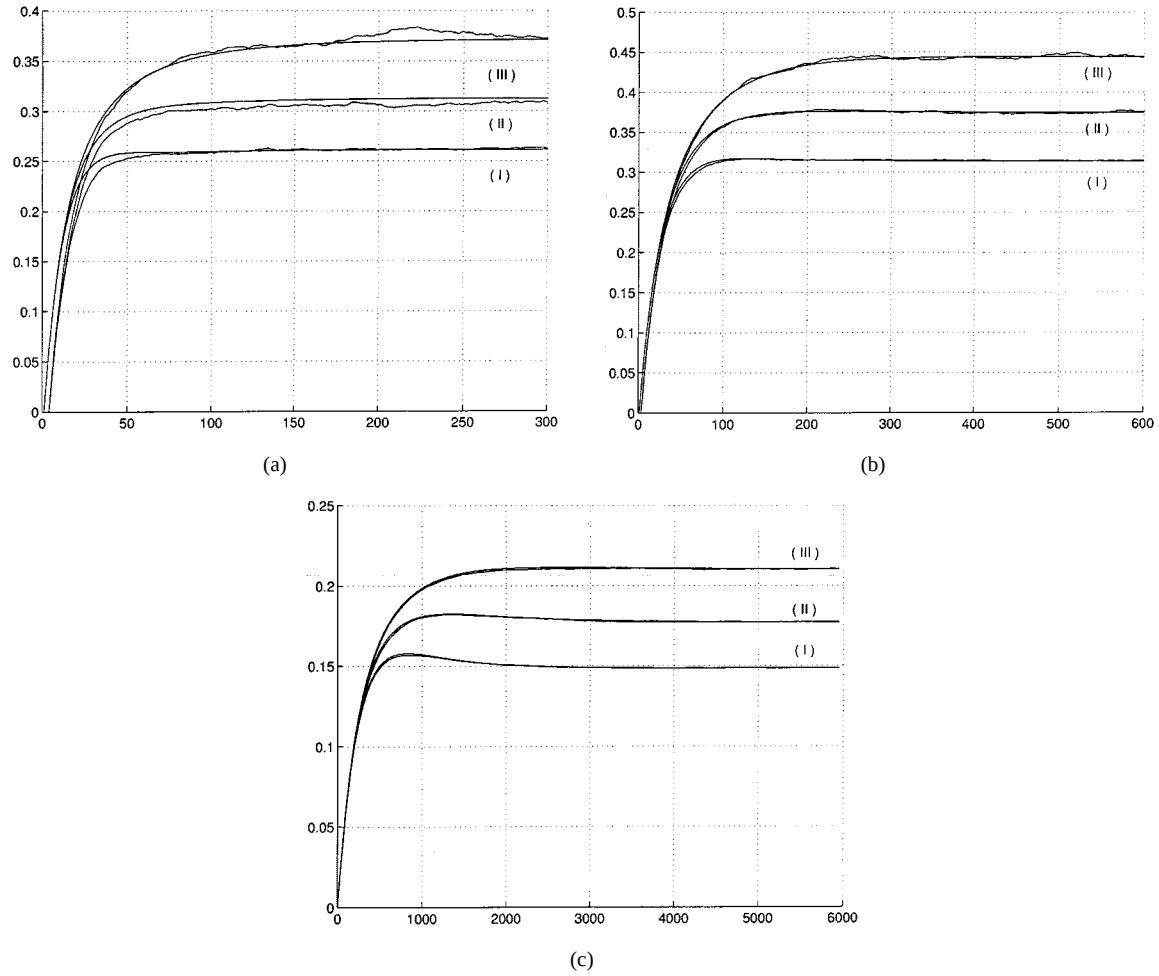

Fig. 2. Example 1: $E\{W(n)\}$ for $\eta^2 = 0.0001$ [curve (I)], 0.3 [curve (II)], and 0.5 [curve (III)]. (a)–(c) Different values of μ. Simulation—ragged curves. Theory—smooth curves. Plots averaged over 1000 runs. (a) $E\{w_4(n)\}$ for $\mu_1 = \mu_{\max}/5 = 0.04$. (b) $E\{w_3(n)\}$ for $\mu_2 = \mu_{\max}/10 = 0.02$. (c) $E\{w_5(n)\}$ for $\mu_3 = \mu_{\max}/100 = 0.002$.

A. Steady-State Mean Weight Behavior

Define

$$\tilde{R}_{ss} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} s_i s_j R_{i-j} \tag{26}$$

$$\tilde{R}_{\hat{s}s} = \sum_{i=0}^{M-1} \sum_{j=0}^{\hat{M}-1} s_i \hat{s}_j R_{i-j} \tag{27}$$

and assume that R_0 and $\tilde{R}_{\hat{s}s}$ are positive definite and that the algorithm converges as $n \to \infty$. Replacing $E\{W(n)\}$ and $E\{W(n+1)\}$ with $W_\infty = \lim_{n\to\infty} E\{W(n)\}$ in (10) yields, after simple algebraic manipulation

$$W_\infty = \tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o \sqrt{\frac{1}{\sigma^2} W_\infty^T \tilde{R}_{ss} W_\infty + 1} \tag{28}$$

where $\tilde{R}_{ss}$ is the autocorrelation matrix of the reference signal $x(n)$ filtered by S. Thus, $\tilde{R}_{ss}$ is positive definite for R_0 positive definite,[7] and $W_\infty^T \tilde{R}_{ss} W_\infty \geq 0$. With these considerations, it is clear from (28) that $W_\infty = a\tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o$, where $a \in \mathbb{R}^+$.

[7]Note that $\tilde{R}_{ss}$ needs to be semi-definite positive for $\sigma^2 > 0$.

Substituting $a\tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o$ for W_∞ in (28) and solving for $a \in \mathbb{R}^+$ yields

$$W_\infty = \lim_{n\to\infty} E\{W(n)\} = \frac{1}{\sqrt{1-\eta^2}} \tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o \tag{29}$$

where η^2 is given by

$$\eta^2 = \frac{1}{\sigma^2} W^{oT} \tilde{R}_{\hat{s}}^T \left(\tilde{R}_{\hat{s}s}^T\right)^{-1} \tilde{R}_{ss} \tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o \tag{30}$$

and corresponds to the system's degree of nonlinearity defined in (2). This can be verified as follows. It is easy to show that for the converged linear case[8]

$$E\left\{y_s^2(n)\right\}\Big|_{\substack{n\to\infty \\ \sigma^2\to\infty}} = W_{\infty_{lin}}^T \tilde{R}_{ss} W_{\infty_{lin}} \tag{31}$$

where $W_{\infty_{lin}} = \tilde{R}_{\hat{s}s}^{-1} \tilde{R}_{\hat{s}} W^o$ is obtained from (29) for $\sigma \to \infty$ ($\eta^2 = 0$, which is the linear case).

Note that the steady-state mean weight vector is not a scaled version of W^o, as is true for the LMS algorithm with a nonlinearity (equivalent to $S = \hat{S} = 1$ in this case) [10]. As expected,

[8]After convergence, $W(n) = W_\infty$ and the order of S and W_∞ can be exchanged.

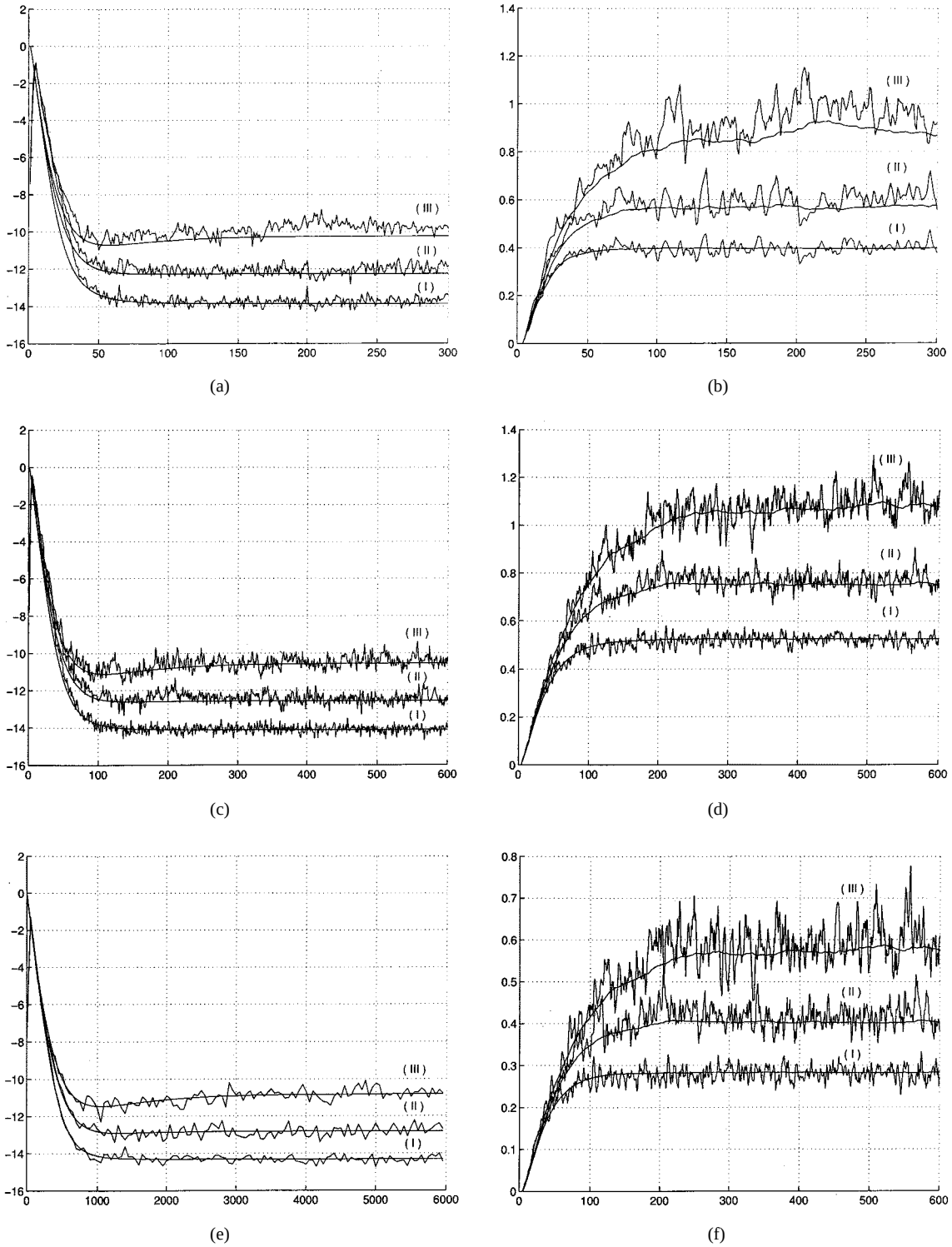

Fig. 3. Example 1: Left column: MSE for $\eta^2 = 0.0001$ [curve (I)], 0.3 [curve (II)], and 0.5 [curve (III)]. Simulation—ragged curves. Theory—smooth curves. Right column: Verification of $E\{W^T(n-i)X(n-i)X^T(n-j)W(n-j)\} \approx E\{W^T(n-i)R_{j-i}W(n-j)\}$ for $\eta^2 = 0.0001$, 0.1, and 0.5. Ragged curves: Actual value. Smooth curves: Approximation. All plots averaged over 1000 runs. (a) MSE for $\mu_1 = \mu_{\max}/5 = 0.04$. (b) $\mu_1 = \mu_{\max}/5 = 0.04$. $i = 0$, $j = 1$. (c) MSE for $\mu_2 = \mu_{\max}/10 = 0.02$. (d) $\mu_2 = \mu_{\max}/10 = 0.02$. $i = 0$, $j = 0$. (e) MSE for $\mu_3 = \mu_{\max}/100 = 0.002$. (f) $\mu_2 = \mu_{\max}/10 = 0.02$. $i = 0$, $j = 2$.

the attainable cancellation level depends on the degree of nonlinearity, the secondary path response S, and the quality of its estimate $\hat{S}$. These effects cannot be reduced by using a smaller μ. As $\sigma^2 \to \infty$ (toward the linear case), $\eta^2 \to 0$, and (29) reduces to the steady-state mean weight vector for the linear case [11, Eqs. (10) and (18)]. On the other hand, $E\{W(n)\}$ is unbounded as $\eta^2 \to 1$, and (29) has no stationary points for $\eta^2 \geq 1$. From (2), $\eta^2 = 1$ sets a power threshold above which the adaptive branch cannot cancel the desired signal $d(n)$. Hence, the adaptive filter gain is never sufficient to overcome the nonlinear saturation.

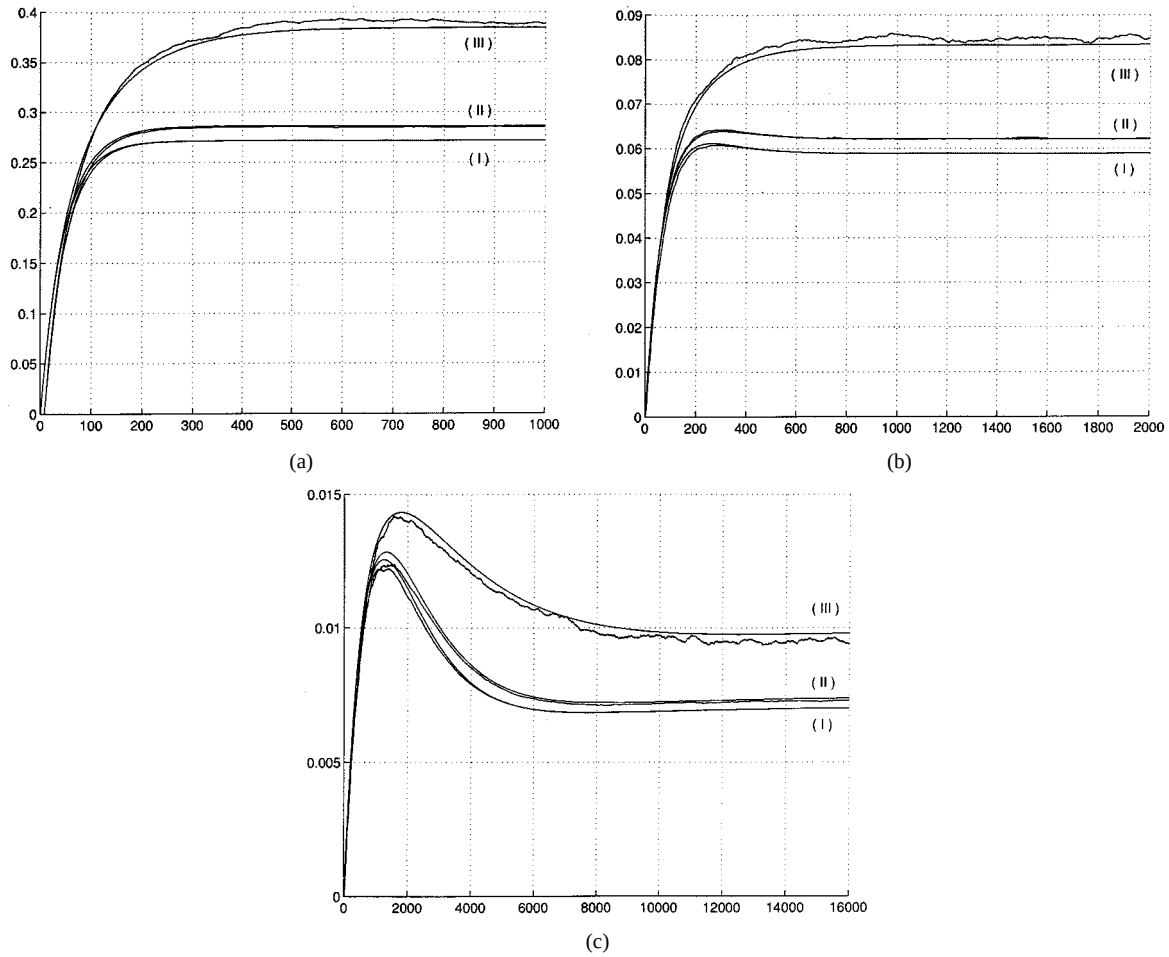

Fig. 4. Example 2: $E\{W(n)\}$ for $\eta^2 = 0.0001$ [curve (I)], 0.1 [curve (II)], and 0.5 [curve (III)]. (a)–(c) Different values of μ. Simulation—ragged curves. Theory—smooth curves. (a) $E\{w_8(n)\}$ for $\mu_1 = \mu_{\max}/5 = 0.01$. (b) $E\{w_2(n)\}$ for $\mu_2 = \mu_{\max}/10 = 0.005$. (c) $E\{w_{14}(n)\}$ for $\mu_3 = \mu_{\max}/100 = 0.0005$.

B. Steady-State MSE

An approximate expression for the steady-state MSE behavior can be determined by replacing $W(n)$ with the steady-state mean weight vector expression (29) and using the definitions of $\tilde{R}_{ss}$, $\tilde{R}_{\hat{s}s}$, and $\tilde{R}_{\hat{s}}$ in (16). After simple algebraic manipulations

$$\begin{aligned}\lim_{n\to\infty} \xi(n) &= W^{o^T}\tilde{R}_{\hat{s}}^T\left(\tilde{R}_{\hat{s}s}^T\right)^{-1}\tilde{R}_{ss}\tilde{R}_{\hat{s}s}^{-1}\tilde{R}_{\hat{s}}W^o\left[\frac{1}{\eta^2}\sin^{-1}(\eta^2)\right] \\ &\quad - 2W^{o^T}\tilde{R}_s^T\tilde{R}_{\hat{s}s}^{-1}\tilde{R}_{\hat{s}}W^o + W^{o^T}R_0W^o + \sigma_z^2. \end{aligned} \tag{32}$$

It is easy to verify that for $\hat{S} = S$ and $\eta^2 \to 0$, (32) tends to [11, Eq. (11)]

$$\xi_{\min} = \sigma_z^2 + W^{o^T}\left\{R_0 - \tilde{R}_s^T\tilde{R}_{ss}^{-1}\tilde{R}_s\right\}W^o \tag{33}$$

which is the expression for the minimum MSE in the linear case.

IV. Simulation Results

This section presents simulation results in support of the theoretical models. Representative plots have been selected from a large number of results.

A. Example 1

Consider $W^o = [0.4130\,0.4627\,0.4803\,0.4627\,0.4130]^T$, $W^{o^T}W^o = 1$, $x(n)$ white with variance $\sigma_x^2 = 1$, measurement noise $z(n)$ with $\sigma_z^2 = 10^{-6}$, and perfect secondary path estimation with $\hat{S} = S = [0.9325\,0.2798\,0.1865\,0.0933\,0.0933]^T$. Simulations are presented for three step sizes (normalized with respect to the linear FXLMS stability limit). The stability limit $\mu_{\max} \approx 0.2$ has been determined by simulation. Step sizes $\mu_1 = \mu_{\max}/5$, $\mu_2 = \mu_{\max}/10$ and $\mu_3 = \mu_{\max}/100$ have been used to evaluate the models for large, moderate, and small μ. In addition, $\eta^2 = 0.0001$, 0.3, and 0.5 have been selected to illustrate the model accuracy for small, moderate, and large degrees of nonlinearity. Fig. 2(a)–(c) compare the simulated mean weight behavior with the analytical predictions using (10), (19), (22), and (25). Each plot presents the results for $\eta^2 = 0.0001$, 0.3, and 0.5 and a single μ averaged over 1000 realizations. The vector components were selected at random. The remaining components have similar behavior. The analytical model is accurate even for relatively large step sizes. The steady-state mean weight behavior, which is predicted by (29), is very accurate, even for the large μ in Fig. 2(a). The predicted steady-state values for $w_4(n)$ by (29) are 0.2614, 0.3124, and 0.3697. Note that the weight fluctuations increase

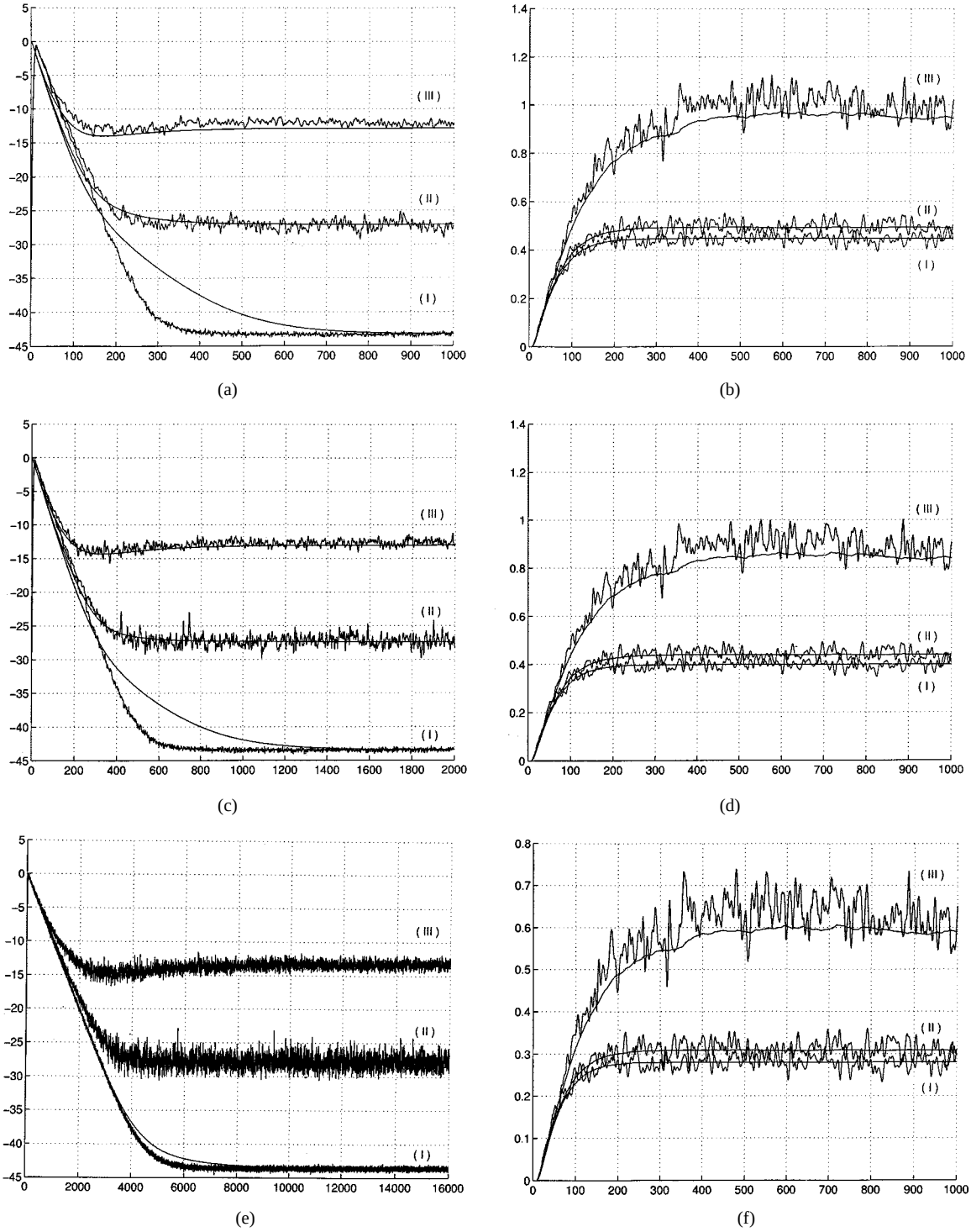

Fig. 5. Example 2: Left column: MSE for $\eta^2 = 0.0001$ [curve (I)], 0.1 [curve (II)], and 0.5 [curve (III)]. Simulation—ragged curves. Theory—smooth curves. Right column: Verification of $E\{W^T(n-i)X(n-i)X^T(n-j)W(n-j)\} \approx E\{W^T(n-i)R_{j-i}W(n-j)\}$ for $\eta^2 = 0.0001$ [curve (I)], 0.1 [curve (II)], and 0.5 [curve (III)]. Ragged curves: Actual value. Smooth curves: Approximation. All plots averaged over 1000 runs. (a) MSE for $\mu_1 = \mu_{\max}/5 = 0.01$. (b) $\mu_1 = \mu_{\max}/5 = 0.01$. $i = 0, j = 0$. (c) MSE for $\mu_2 = \mu_{\max}/10 = 0.005$. (d) $\mu_1 = \mu_{\max}/5 = 0.01$. $i = 0, j = 2$. (e) MSE for $\mu_3 = \mu_{\max}/100 = 0.0005$. (f) $\mu_1 = \mu_{\max}/5 = 0.01$. $i = 0, j = 4$.

with η^2. This increase is probably due to saturation, which clips the adaptive filter output for larger η^2. Clipping in turn increases the error signal and the weight update.

Fig. 3(a), (c), and (e) show the simulated MSE and the theoretical MSE using (16), (10), (19), (22), and (25). Each figure shows curves for $\eta^2 = 0.0001$, 0.3, and 0.5. Plots are shown for different step sizes. All plots were obtained by averaging 1000 runs. The analytical model and the simulations are in close agreement in all cases, even for relatively large $\mu = \mu_1$. The steady-state MSE [predicted by the closed form in (32)] are

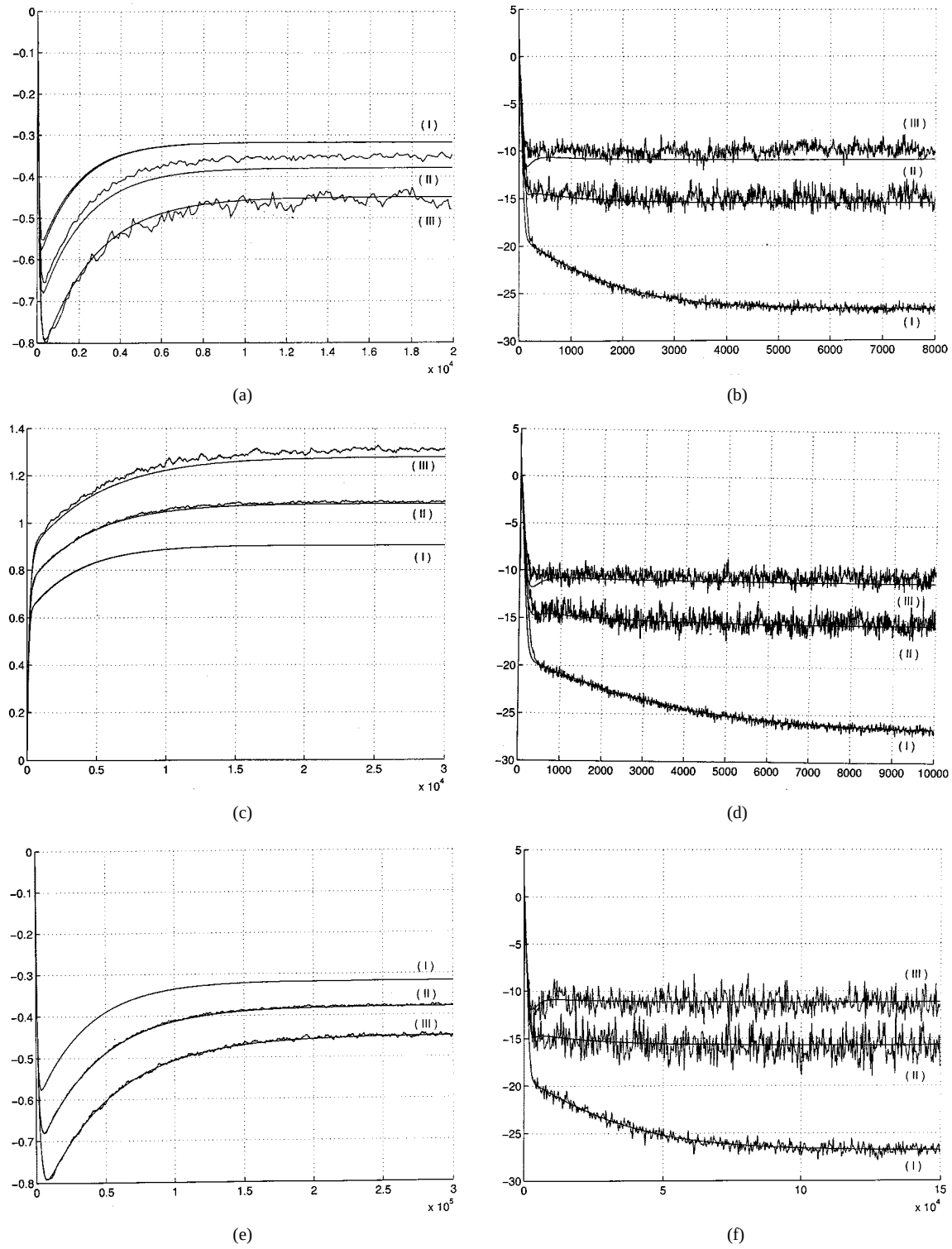

Fig. 6. Example 3. Left column: $E\{W(n)\}$ for $\eta^2 = 0.0001$ [curve (I)], 0.3 [curve (II)], and 0.5 [curve (III)]. Right column: MSE for $\eta^2 = 0.0001$ [curve (I)], 0.3 [curve (II)], and 0.5 [curve (III)]. Simulation—ragged curves. Theory—smooth curves. All plots averaged over 1000 runs. (a) $E\{w_3(n)\}$ for $\mu_1 = \mu_{\max}/5 = 0.012$. (b) MSE for $\mu_1 = \mu_{\max}/5 = 0.012$. (c) $E\{w_1(n)\}$ for $\mu_2 = \mu_{\max}/10 = 0.006$. (d) MSE for $\mu_2 = \mu_{\max}/10 = 0.006$. (e) $E\{w_3(n)\}$ for $\mu_3 = \mu_{\max}/100 = 0.0006$. (f) MSE for $\mu_3 = \mu_{\max}/100 = 0.0006$.

-14.34 dB, -12.85 dB, and -10.85 dB. These values are accurate only for small μ [Fig. 3(e)] since they do not account for the weight fluctuations. The recursive model should be run to accurately predict the steady-state behavior for moderate and large μ. Fig. 3(b), (d), and (f) verify the accuracy of assumption A1 for different i, j, μ, and η^2.

B. Example 2

This example repeats Example 1 for a longer impulse response W^o and for an imperfect estimate of the secondary path. Consider $W^o = [0.0156\ 0.0598\ 0.1260\ 0.2041\ 0.2822\ 0.3485\ 0.3927\ 0.4083\ 0.3927\ 0.3485\ 0.2822\ 0.2041\ 0.1260\ 0.0598\ 0.0156]^T$, $W^{oT}W^o = 1$, $S =$

$[0.9356\ 0.2807\ 0.1871\ 0.0936\ 0.0468]^T$, $S^TS = 1$, $\hat{S} = [0.8922\,0.3965\ 0.1487\ 0.1487\ 0.0496]^T$, $\hat{S}^T\hat{S} = 1$, $x(n)$ white with $\sigma_x^2 = 1$ and $\sigma_z^2 = 10^{-6}$. The stability limit $\mu_{\max} = 0.05$ was determined by simulation. The parameters were $\mu_1 = \mu_{\max}/5$, $\mu_2 = \mu_{\max}/10$, $\mu_3 = \mu_{\max}/100$, and $\eta^2 = 0.0001$, 0.1, and 0.5.

Figs. 4 and 5 verify the analytical model using recursions (10), (16), (19), (22), and (25). The right column of Fig. 5 validates assumption A1 for different delays and $\mu = \mu_1$. Fig. 5(a) and (c) suggest that the model can deviate from the simulation for large step sizes and very small degrees of nonlinearity. This mismatch is probably due to some of the stochastic approximations being based on slow algorithm learning. Fig. 5(e) shows that the mismatch is minimal for small μ used in most practical applications (see [18] for instance). The model is accurate for the initial transient phase (cancellation to -30 dB, compatible with most practical applications) and in steady-state, even for large μ. The excess MSE (due to the nonlinearity) increases as the system deviates from the linear case. Then, the inaccuracies of the analytical model can be neglected even for large step sizes [see curves (II) and (III) in Fig. 5(a), (c), and (e)]. The significant impact of the nonlinearity on the cancellation level (steady-state MSE) is illustrated by this example. Clearly, the nonlinear model is required to accurately predict this cancelation level.

C. Example 3

This example verifies the model accuracy for correlated inputs. $x(n)$ is an autoregressive process with $\sigma_x^2 = 1$ that is obtained by passing a white noise $u(n)$ with variance $\sigma_u^2 = 0.0965$ through the filter with attenuation given by $A(z) = 1 - 0.195z^{-1} + 0.95z^{-2}$. The eigenvalue spread of R_0 is equal to 39.82 [2]. $W^o = [0.7756\,0.5171\,-0.3620]^T$, $S = [0.8944\,0.4472]^T$, $\hat{S} = [0.9701\,0.2425]^T$ (imperfect secondary path estimation). $\mu_{\max} = 0.06$ (experimentally obtained for the linear case). The parameters used were again $\mu_1 = \mu_{\max}/5$, $\mu_2 = \mu_{\max}/10$, $\mu_3 = \mu_{\max}/100$, and $\eta^2 = 0.0001$, 0.3, and 0.5. Fig. 6 shows the theoretical and simulated results. Fig. 6(a), (c), and (e) compare the simulations and the mean weight behavior predicted by the model. Fig. 6(b), (d), and (f) show the results for the MSE. Note that the model accurately predicts the simulation behavior.

V. Conclusion

This paper has presented a statistical analysis of the filtered-X LMS (FXLMS) algorithm with a nonlinear secondary path. This structure can model nonlinear effects in active noise and active vibration control systems when transducers are driven by large amplitude signals. Deterministic nonlinear recursions were derived for Gaussian inputs for the transient mean weight, mean square error, and cross-covariance matrix of the adaptive weight vector at different times. The new results generalize previous results obtained for linear secondary paths. The cross-covariance results provide improved steady-state predictions (as compared with previous results) for moderate to large step sizes. Closed-form expressions have been derived for small step sizes for the steady-state mean weight and MSE. Monte Carlo simulations displayed excellent agreement with the theoretical predictions for both small and large step sizes. This agreement provides strong support for the approximations used to derive the theoretical model.

Appendix A
Derivation of $E\{W(n-i+1)W^T(n-j+1)|\Psi\}$

Multiplying delayed versions of (6) and taking the expected value conditioned in Ψ yields (34), shown at the bottom of the next page.

The expected values in (34) are now determined.

Expression 1:

$$\begin{aligned} &E\left\{X(n-j)X_f^T(n-j)|\Psi\right\} \\ &\quad = \sum_{k=0}^{\hat{M}-1} \hat{s}_k E\left\{X(n-j)X^T(n-j-k)|\Psi\right\} \\ &\quad = \sum_{k=0}^{\hat{M}-1} \hat{s}_k R_k = \tilde{R}_{\hat{s}}^T. \end{aligned} \tag{35}$$

Expression 2: Assuming $X(n-p)$ statistically independent of Ψ, $\forall p$

$$\begin{aligned} &E\left\{z(n-j)X_f^T(n-j)|\Psi\right\} \\ &\qquad = E\left\{z(n-j)\right\} E\left\{X_f^T(n-j)|\Psi\right\} = 0. \end{aligned} \tag{36}$$

Expression 3: The delayed version of the expression has already been evaluated for (8).

Expression 4: Assuming $X(n-p)$ statistically independent of Ψ, $\forall p$

$$\begin{aligned} &E\left\{z(n-i)z(n-j)X_f(n-i)X_f^T(n-j)|\Psi\right\} \\ &\quad = E\left\{z(n-i)z(n-j)\right\} E\left\{X_f(n-i)X_f^T(n-j)\right\} \\ &\quad = \sigma_z^2\,\delta(i-j) \sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1} \hat{s}_p\hat{s}_q E\left\{X(n-i-p)X^T(n-j-q)\right\} \\ &\quad = \sigma_z^2\,\delta(i-j) \sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1} \hat{s}_p\hat{s}_q R_{j-i+q-p}. \end{aligned} \tag{37}$$

Expression 5:

$$\begin{aligned} &E\left\{z(n-j)W^{o^T}X(n-i)X_f(n-i)X_f^T(n-j)|\Psi\right\} \\ &\quad = E\{z(n-j)\}E\left\{W^{o^T}X(n-i)X_f(n-i)\right. \\ &\qquad \left.\cdot X_f^T(n-j)|\Psi\right\} = 0. \end{aligned} \tag{38}$$

Expression 6: Using the moment-factoring theorem for Gaussian variates [2], it can be easily shown that

$$E\left\{X_f(n-i)X^T(n-i)W^oW^{o^T}X(n-j)X_f^T(n-j)|\Psi\right\} = \sum_{p=0}^{\hat{M}-1}\hat{s}_pR_{-p}W^oW^{o^T}\sum_{q=0}^{\hat{M}-1}\hat{s}_qR_q + \sum_{p=0}^{\hat{M}-1}\hat{s}_pR_{j-i-p}W^oW^{o^T}\sum_{q=0}^{\hat{M}-1}\hat{s}_qR_{j-i+q} + W^{o^T}R_{j-i}W^o\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\hat{s}_p\hat{s}_qR_{j-i+q-p}. \quad (39)$$

$$\begin{aligned}
&E\{W(n-i+1)W^T(n-j+1)|\Psi\}\\
&= W(n-i)W^T(n-j) + \mu W(n-i)W^{o^T}\overbrace{E\{X(n-j)X_f^T(n-j)|\Psi\}}^{(1)}\\
&\quad + \mu\left[W(n-j)W^{o^T}E\{X(n-i)X_f^T(n-i)|\Psi\}\right]^T + \mu W(n-i)\overbrace{E\{z(n-j)X_f^T(n-j)|\Psi\}}^{(2)}\\
&\quad + \mu\left[W(n-j)E\{z(n-i)X_f^T(n-i)|\Psi\}\right]^T\\
&\quad - \mu\left[\overbrace{E\left\{g\left[\sum_{k=0}^{M-1}s_kW^T(n-i-k)X(n-i-k)\right]X_f(n-i)\middle|\Psi\right\}}^{(3)}W^T(n-j)\right]\\
&\quad - \mu\left[E\left\{g\left[\sum_{l=0}^{M-1}s_lW^T(n-j-l)X(n-j-l)\right]X_f(n-j)\middle|\Psi\right\}W^T(n-i)\right]^T\\
&\quad + \mu^2\overbrace{E\{z(n-i)z(n-j)X_f(n-i)X_f^T(n-j)|\Psi\}}^{(4)} + \mu^2\overbrace{E\{z(n-j)W^{o^T}X(n-i)X_f(n-i)X_f^T(n-j)|\Psi\}}^{(5)}\\
&\quad + \mu^2\left[E\{z(n-i)W^{o^T}X(n-j)X_f(n-j)X_f^T(n-i)|\Psi\}\right]^T\\
&\quad + \mu^2\overbrace{E\{X_f(n-i)X^T(n-i)W^oW^{o^T}X(n-j)X_f^T(n-j)|\Psi\}}^{(6)}\\
&\quad - \mu^2\overbrace{E\left\{g\left[\sum_{k=0}^{M-1}s_kW^T(n-i-k)X(n-i-k)\right]W^{o^T}X(n-j)X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}}^{(7)}\\
&\quad - \mu^2E\left\{g\left[\sum_{l=0}^{M-1}s_lW^T(n-j-l)X(n-j-l)\right]W^{o^T}X(n-i)X_f(n-j)X_f^T(n-i)\middle|\Psi\right\}^T\\
&\quad - \mu^2\overbrace{E\left\{z(n-i)g\left[\sum_{l=0}^{M-1}s_lW^T(n-j-l)X(n-j-l)\right]X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}}^{(8)}\\
&\quad - \mu^2E\left\{z(n-j)g\left[\sum_{k=0}^{M-1}s_kW^T(n-i-k)X(n-i-k)\right]X_f(n-j)X_f^T(n-i)\middle|\Psi\right\}^T\\
&\quad + \mu^2\overbrace{E\left\{g\left[\sum_{k=0}^{M-1}s_kW^T(n-i-k)X(n-i-k)\right]\cdot g\left[\sum_{l=0}^{M-1}s_lW^T(n-j-l)X(n-j-l)\right]X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}}^{(9)} \quad (34)
\end{aligned}$$

$$E\left\{g\left[\sum_{k=0}^{M-1} s_k W^T(n-i-k)X(n-i-k)\right] W^{o^T} X(n-j)X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}$$

$$= -\frac{1}{\sigma^2\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q W^T(n-i-p)R_{q-p}W(n-i-q)+1\right)^{3/2}}$$

$$\cdot \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1} s_r s_k \hat{s}_l s_p \hat{s}_q W^T(n-i-r)R_{j-i-r}W^o R_{k-l}W(n-i-k)W^T(n-i-p)R_{j-i+q-p}$$

$$+ \frac{1}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} s_p s_q W^T(n-i-p)R_{q-p}W(n-i-q)+1}}$$

$$\cdot \left[\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1} \hat{s}_p \hat{s}_q s_r W^T(n-i-r)R_{j-i-r}W^o R_{j-i+q-p} + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1} s_p \hat{s}_q R_{p-q}W(n-i-p)W^{o^T}\tilde{R}_{\hat{s}}^T\right.$$

$$\left. + \sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{\hat{M}-1} s_p \hat{s}_q \hat{s}_r R_{j-i-r}W^o W^T(n-i-p)R_{j-i+q-p}\right]. \quad (40)$$

Expression 7: Using the same approach used to derive [17, Eq. A6] and [10, Eq. 61], after laborious calculations, we have (40), shown at the top of the page.

Expression 8:

$$E\left\{z(n-i)g\left[\sum_{l=0}^{M-1} s_l W^T(n-j-l)X(n-j-l)\right]\right.$$
$$\left.\cdot X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}$$
$$= E\{z(n-i)\}E\left\{g\left[\sum_{l=0}^{M-1} s_l W^T(n-j-l)X(n-j-l)\right]\right.$$
$$\left.\cdot\ X_f(n-i)X_f^T(n-j)\middle|\Psi\right\} = 0. \quad (41)$$

Expression 9: An expression for this expected value can be derived using the same approaches used to derive [17, Eq. A13] and [10, Eq. 69]. However, this involved expression is not necessary for the purposes of this paper, as is explained in the main text. Thus, this term will be represented by the short functional notation

$$Q\left(S, \hat{S}, R_{k-\ell}, n, i, j, \Psi\right).$$

Combining all terms of (34), we have (42), shown on the next page, where $\tilde{R}_{\hat{s}} = \sum_{j=0}^{\hat{M}-1} \hat{s}_j R_{-j}$, and

$$Q\left(S, \hat{S}, R_{k-\ell}, n, i, j, \Psi\right)$$
$$= E\left\{g\left[\sum_{k=0}^{M-1} s_k W^T(n-i-k)X(n-i-k)\right]\right.$$
$$\cdot g\left[\sum_{l=0}^{M-1} s_l W^T(n-j-l)X(n-j-l)\right]$$
$$\left.\cdot\ X_f(n-i)X_f^T(n-j)\middle|\Psi\right\}. \quad (43)$$

Appendix B
Implementation of the Analytical Model

1. Initialization:

1.1) Let $M_0 = \max\{M, \hat{M}\}$.

1.2) Determine R_c for $-M_0 \le c \le M_0$, $\tilde{R}_{\hat{s}} = \sum_{j=0}^{\hat{M}-1} \hat{s}_j R_{-j}$, and $\tilde{R}_{\hat{s}\hat{s}} = \sum_{i=0}^{\hat{M}-1}\sum_{j=0}^{\hat{M}-1} \hat{s}_i \hat{s}_j R_{i-j}$.

1.3) Make

$$\begin{cases} E\{W(0)\} = 0, \\ K_{0,k}(n) = 0, & 0 \le k \le M_0 - 1 \text{ and } n \le 0. \\ \xi(0) = W^{o^T} R_0 W^o + \sigma_z^2 \end{cases}$$

2. Iterations:

2.1) Evaluate $K_{0,0}(n+1)$ using (10), (19), and (22).

2.2) Evaluate $K_{0,k}$ for $1 \le k \le M_0 - 1$ using (10), (19), and (25).

2.3) Evaluate $E\{W(n+1)\}$ using (10) and (19).

2.4) Evaluate $\xi(n+1)$ using (10), (16), and (19).

2.5) Return to step 2.1).

$$
\begin{aligned}
&E\left\{W(n-i+1)W^T(n-j+1)|\Psi\right\}\\
&\quad = W(n-i)W^T(n-j)+\mu W(n-i)W^{o^T}\tilde{R}_{\hat{s}}^T+\mu\tilde{R}_{\hat{s}}W^oW^T(n-j)\\
&\qquad -\frac{\mu\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}W(n-i-p)W^T(n-j)}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-i-p)R_{q-p}W(n-i-q)+1}}\\
&\qquad -\frac{\mu\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qW(n-i)W^T(n-j-p)R_{p-q}^T}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-j-p)R_{q-p}W(n-j-q)+1}}\\
&\qquad +\mu^2\sigma_z^2\delta(i-j)\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\hat{s}_p\hat{s}_qR_{j-i+q-p}+\mu^2\sum_{p=0}^{\hat{M}-1}\hat{s}_pR_{-p}W^oW^{o^T}\sum_{q=0}^{\hat{M}-1}\hat{s}_qR_q\\
&\qquad +\mu^2\sum_{p=0}^{\hat{M}-1}\hat{s}_pR_{j-i-p}W^oW^{o^T}\sum_{q=0}^{\hat{M}-1}\hat{s}_qR_{j-i+q}+\mu^2W^{o^T}R_{j-i}W^o\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\hat{s}_p\hat{s}_qR_{j-i+q-p}\\
&\qquad +\frac{\mu^2}{\sigma^2\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-i-p)R_{q-p}W(n-i-q)+1\right)^{3/2}}\\
&\qquad \cdot\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1}s_rs_k\hat{s}_l\hat{s}_p\hat{s}_q\overbrace{W^T(n-i-r)R_{j-i-r}W^oR_{k-l}W(n-i-k)W^T(n-i-p)R_{j-i+q-p}}^{(A)}\\
&\qquad +\frac{\mu^2}{\sigma^2\left(\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-j-p)R_{q-p}W(n-j-q)+1\right)^{3/2}}\\
&\qquad \cdot\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{k=0}^{M-1}\sum_{l=0}^{\hat{M}-1}\sum_{r=0}^{M-1}s_rs_k\hat{s}_l\hat{s}_p\hat{s}_q\overbrace{W^T(n-j-r)R_{i-j-r}W^oR_{i-j+q-p}^TW(n-j-p)W^T(n-j-k)R_{k-l}^T}^{(B)}\\
&\qquad -\frac{\mu^2}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-i-p)R_{q-p}W(n-i-q)+1}}\\
&\qquad \cdot\left[\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1}\hat{s}_p\hat{s}_qs_rW^T(n-i-r)R_{j-i-r}W^oR_{j-i+q-p}+\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_qR_{p-q}W(n-i-p)W^{o^T}\tilde{R}_{\hat{s}}^T\right.\\
&\qquad\quad \left.+\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{\hat{M}-1}s_p\hat{s}_q\hat{s}_rR_{j-i-r}W^oW^T(n-i-p)R_{j-i+q-p}\right]\\
&\qquad -\frac{\mu^2}{\sqrt{\frac{1}{\sigma^2}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1}s_ps_qW^T(n-j-p)R_{q-p}W(n-j-q)+1}}\\
&\qquad \cdot\left[\sum_{p=0}^{\hat{M}-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{M-1}\hat{s}_p\hat{s}_qs_rW^T(n-j-r)R_{i-j-r}W^oR_{i-j+q-p}^T+\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}s_p\hat{s}_q\tilde{R}_{\hat{s}}W^oW^T(n-j-p)R_{p-q}^T\right.\\
&\qquad\quad \left.+\sum_{p=0}^{M-1}\sum_{q=0}^{\hat{M}-1}\sum_{r=0}^{\hat{M}-1}s_p\hat{s}_q\hat{s}_rR_{i-j+q-p}^TW(n-j-p)W^{o^T}R_{i-j-r}^T\right]+\mu^2Q\left(S,\,\hat{S},\,R_{k-\ell},\,n,\,i,\,j,\,\Psi\right)
\end{aligned}
\tag{42}
$$

REFERENCES

[1] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations*. New York: Wiley, 1996.

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

[3] A. Stenger and W. Kellerman, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, pp. 1747–1760, 2000.

[4] C. Hansen, "Active noise control—From laboratory to industrial implementation," in *Proc. Nat. Conf. Noise Contr. Eng.*, 1997, pp. 3–38.

[5] G. Tao and P. V. Kokotovic, *Adaptive Control of Systems With Actuators and Sensor Nonlinearities*. New York: Wiley, 1996.

[6] W. Klippel, "Dynamic measurements and interpretation of the nonlinear parameters of electrodynamic loudspeakers," *J. Audio Eng. Soc.*, vol. 38, no. 12, pp. 944–955, Dec. 1990.

[7] ——, "Nonlinear large-signal behavior of electrodynamic loudspeakers at low frequencies," *J. Audio Eng. Soc.*, vol. 40, no. 6, pp. 483–496, June 1992.

[8] R. J. Bernhard, P. Davies, and S. W. Kurth, "Effects of nonlinearities on system identification in active noise control systems," in *Proc. Nat. Conf. Noise Contr. Eng.*, 1997, pp. 231–236.

[9] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Statistical analysis of the LMS algorithm with a zero-memory nonlinearity after the adaptive filter," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, paper 1815.

[10] ——, "Stochastic analysis of the LMS algorithm with a saturation nonlinearity following the adaptive filter output," *IEEE Trans. Signal Processing*, vol. 49, pp. 1370–1387, July 2001.

[11] O. J. Tobias, J. C. M. Bermudez, and N. J. Bershad, "Mean weight behavior of the Filtered-X LMS algorithm," *IEEE Trans. Signal Processing*, vol. 48, pp. 1061–1075, Apr. 2000.

[12] O. J. Tobias, J. C. M. Bermudez, N. J. Bershad, and R. Seara, "Second moment analysis of the Filtered-X LMS algorithm," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, paper no. 1584.

[13] O. J. Tobias, J. C. M. Bermudez, R. Seara, and N. Bershad, "An improved model for the second moment behavior of the Filtered-X LMS algorithm," in *Proc. IEEE Adaptive Syst. Signal Process., Commun., Contr. Symp.*, Lake Louise, AB, Canada, Oct. 2000, pp. 337–341.

[14] J. J. Shynk and N. J. Bershad, "Steady-state analysis of a single-layer perceptron based on a system identification model with bias terms," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1030–1042, Sept. 1991.

[15] C. Samson and V. U. Reddy, "Fixed point error analysis of the normalized ladder algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1177–1191, Oct. 1983.

[16] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Statistical analysis of the FXLMS algorithm with a nonlinearity in the secondary path," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 3, Orlando, FL, June 1999, pp. 166–169.

[17] N. J. Bershad, P. Celka, and J. M. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for Gaussian data and noisy input and output measurements," *IEEE Trans. Signal Processing*, vol. 47, pp. 675–689, Mar. 1999.

[18] C. Breining *et al.*, "Acoustic echo control: An application of very-high-order adaptive filters," *IEEE Signal Processing Mag.*, pp. 42–69, July 1999.

Márcio H. Costa received the B.E.E. degree from Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil, in 1991, the M.Sc. degree in biomedical engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brazil, in 1994, and the Dr. degree in electrical engineering from Universidade Federal de Santa Catarina, Florianópolis, Brazil, in 2001.

He joined the Department of Electrical Engineering, Universidade Católica de Pelotas (UCPel), Pelotas, Brazil, in 1994, where he is currently an Associate Professor of electrical engineering and a researcher with the Biomedical Engineering Group. His research interests have involved biomedical signal processing and instrumentation. His present research interests are in discrete-time signal processing, linear and nonlinear adaptive filters, adaptive inverse control, and active noise and vibration control.

José Carlos M. Bermudez (M'85) received the B.E.E. degree from Federal University of Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brazil, in 1978, the M.Sc. degree in electrical engineering from COPPE/UFRJ, in 1981, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 1985.

He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 1985, where he is currently a Professor of electrical engineering. In the winter of 1992, he was a Visiting Researcher with the Department of Electrical Engineering, Concordia University. In 1994, he was a Visiting Researcher with the Department of Electrical and Computer Engineering, University of California, Irvine. His research interests have involved analog signal processing using continuous-time and sampled-data systems. His recent research interests are in digital signal processing, including linear and nonlinear adaptive filtering, active noise and vibration control, acoustic echo cancellation, image processing, and speech processing.

Prof. Bermudez served as an Associate Editor of the IEEE TRANSACTION ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001. He is presently a Member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

Neil J. Bershad (S'60–M'62–SM'81–F'88) received the B.E.E. degree from Rensselaer Polytechnic Institute (RPI), Troy, NY, in 1958, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1960, and the Ph.D. degree in electrical engineering from RPI in 1962.

He joined the Faculty of the School of Engineering, University of California, Irvine, in 1966 and is now an Emeritus Professor of electrical engineering. He was a Visiting Professor of electrical engineering at ENSEEIHT-GAPSE, Tolouse, France, from 1994 to 1998, at the Signal Processing Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, from 1997 to 1999, and at the University of Edinburgh, Edinburgh, U.K., in 1999. His research interests have involved stochastic systems modeling and analysis. His recent interests are in the area of stochastic analysis of adaptive filters. He has published a significant number of papers on the analysis of the stochastic behavior of various configurations of the LMS adaptive filter. His present research interests include the statistical learning behavior of adaptive filter structures for nonlinear signal processing, neural networks when viewed as nonlinear adaptive filters, and active acoustic noise cancelation.

Dr. Bershad has served as an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of phase-locked loops and synchronization. More recently, he was an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING in the area of adaptive filtering.