Stochastic Analysis of the LMS Algorithm with a Saturation Nonlinearity Following the Adaptive Filter Output

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

***Abstract*—This paper presents a statistical analysis of the least mean square (LMS) algorithm with a zero-memory scaled error function nonlinearity following the adaptive filter output. This structure models saturation effects in active noise and active vibration control systems when the acoustic transducers are driven by large amplitude signals. The problem is first defined as a nonlinear signal estimation problem and the mean-square error (MSE) performance surface is studied. Analytical expressions are obtained for the optimum weight vector and the minimum achievable MSE as functions of the saturation. These results are useful for adaptive algorithm design and evaluation. The LMS algorithm behavior with saturation is analyzed for Gaussian inputs and slow adaptation. Deterministic nonlinear recursions are obtained for the time-varying mean weight and MSE behavior. Simplified results are derived for white inputs and small step sizes. Monte Carlo simulations display excellent agreement with the theoretical predictions, even for relatively large step sizes. The new analytical results accurately predict the effect of saturation on the LMS 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]–[3]. Most adaptive system analyses assume nonlinear effects can be neglected and model both the unknown system and the adaptive path as linear with memory. Linearity simplifies the mathematical problem and often permits a detailed system analyses in many important practical circumstances. However, more sophisticated models must be used when nonlinear effects are significant to the system behavior (i.e., amplifier saturation).

Manuscript received April 12, 2000; revised March 16, 2001. This work was supported in part by the Brazilian Ministry of Education (CAPES) under Grant PICDT 0120/97-9 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. Kristine L. Bell.

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(01)05180-7.

Linear adaptive cancellation paths are the natural design choice in linear system identification. However, numerous practical adaptive systems have significant intrinsic nonlinearities in the cancellation path. Such nonlinearities are unavoidable and their effects on the overall adaptive system behavior must be considered in a design situation. Important application examples are active noise control (ANC) and active vibration control (AVC) systems. ANC and AVC systems include acoustical/mechanical paths. Signal converters (A/D and D/A), power amplifiers, and transducers (speakers or actuators) transform digital electrical signals into analog electrical or mechanical signals for proper cancellation [1]. System or secondary path nonlinearities[1] can become important nonideal effects in ANC and AVC systems [4], [5]. The nonlinearity can be caused by overdriving the electronic circuitry or the speakers/transducers in the secondary path, for example.

In [5], Bernhard *et al.* briefly discussed nonlinear effects in ANC systems, but no adaptive algorithm behavior analysis was presented. In [4], Snyder and Tanaka propose modeling a nonlinear primary path with a neural network nonlinear controller in the AVC system. Again, no analysis was presented for algorithm behavior. Most practical ANC and AVC systems contain nonlinearities in the secondary path. Therefore, it is of great interest to determine the effect of such nonlinearities on the adaptive algorithm. Such analysis is unavailable in the open literature. Several researchers have studied the statistical behavior of the LMS algorithm with nonlinearities applied to the correlation multiplier. Representative examples are [6]–[15]. These results cannot be modified to explain algorithm behavior with a nonlinearity at the adaptive filter output.

This paper investigates the statistical behavior of the system in Fig. 1. The function $g(y)$ is a zero-memory saturation nonlinearity. Stochastic analysis of this system can provide important insights into nonlinear secondary path effects upon ANC and AVC system behavior. Neural networks can be viewed as adaptive filters with output nonlinearities during the learning phase. Thus, the results presented here may also be useful for studying the statistical behavior of neural networks.

Fig. 1 is analyzed first as estimation of a sequence $d(n)$ from a nonlinear function of the reference signal $x(n)$. The mean square error (MSE) performance surface properties are determined as functions of the system's degree of nonlinearity (defined below). The MSE surface is shown to deform due to the

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

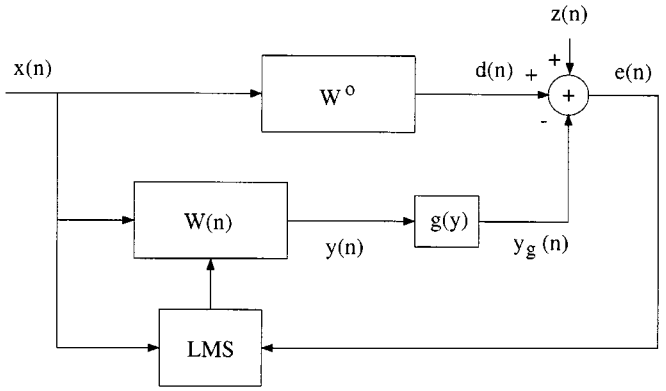


Fig. 1. Block diagram of the nonlinear adaptive system.

nonlinearity but remains unimodal. The optimum weight vector is a scaled version of the Wiener weight for the linear case.

Deterministic nonlinear recursions are derived for the mean weight and mean square error (MSE) behaviors of the LMS adaptive algorithm for Gaussian inputs and slow adaptation. The LMS algorithm introduces a multiplicative bias in the converged mean weight vector (compared with the optimum solution). The degree of nonlinearity is shown to affect the algorithm behavior and the achievable level of cancellation. Monte Carlo simulations display excellent agreement with the theoretical predictions.

II. Analysis of the MSE Surface

Consider initially the nonadaptive system shown in Fig. 2. This block diagram corresponds to a nonlinear mean-square estimation problem [16, Sec. 7–5]. The sequence $d(n)$ is estimated in the mean square sense by a nonlinear function of the reference signal $x(n)$. The properties of the MSE surface as a function of the system's degree of nonlinearity is studied here for Gaussian inputs.

A. Analysis Model

The notation for Fig. 2 is as follows.

$W^o = [w_0^o\ w_1^o \cdots w_{N-1}^o]^{\mathrm{T}}$	Response of the unknown system.
$W = [w_0\ w_1 \cdots w_{N-1}]^{\mathrm{T}}$	Linear filter weight vector.
$x(n)$	Reference signal.
$X(n) = [x(n)\,x(n-1)\cdots x(n-N+1)]^{\mathrm{T}}$	Observed input data vector.
$d(n)$	Primary signal.
$z(n)$	Measurement noise.
$y(n)$	Output of the linear filter.
$g(y)$	Saturation nonlinearity.
$y_g(n)$	Nonlinearity output.

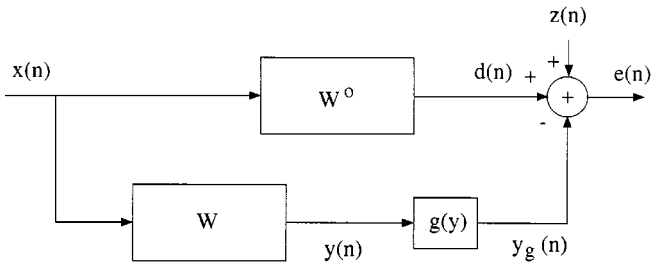


Fig. 2. Nonlinear optimal filtering problem.

$x(n)$ 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 uncorrelated with any other signal. The saturation nonlinearity is modeled by the scaled error function

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

The system's degree of nonlinearity is controlled by the parameter σ^2 in (1) and is defined as

$$\eta^2 = \frac{\pi}{2}\frac{\sigma_d^2}{\max\left\{\sigma_{y_g}^2\right\}} = \frac{1}{\sigma^2} W^{o^{\mathrm{T}}} R_{xx} W^o \tag{2}$$

where

$R_{xx} = E\{X(n)X^{\mathrm{T}}(n)\}$	autocorrelation matrix of the input vector;
σ_d^2	variance of $d(n)$;
$\max\left\{\sigma_{y_g}^2\right\} = (\pi/2)\sigma^2$	maximum variance of $y_g(n)$ obtained by taking the limit of (1) as $y \to \infty$.

Equation (2) expresses the ratio of the power in $d(n)$ (ideal output of W for the linear case) to the maximum available power in $y_g(n)$, which is the cancelling signal. 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, the behavior of $g(y)$ can be varied between that of a linear device and that of a hard limiter by changing σ^2. The effects of very large nonlinearities ($\sigma \to 0$) can be studied by scaling $g(y)$ by a constant such as A/σ, $A \in \mathbb{R}^+$. This artifice avoids the attenuation factor σ in the limit as sigma approaches zero. This paper studies the algorithm behavior for $g(y)$ in (1) that models the degrees of nonlinearity of most interest in practical applications. Results for very large degrees of nonlinearity can easily be obtained from the results presented here by carrying the effect of A/σ throughout the derivations[2].

B. MSE Performance Surface

The error signal in Fig. 2 is given by

$$\begin{aligned} e(n) &= d(n) + z(n) - y_g(n) \\ &= W^{o^{\mathrm{T}}} X(n) + z(n) - g\left[W^{\mathrm{T}} X(n)\right]. \end{aligned} \tag{3}$$

[2]In this case, $\max\left\{\sigma_{y_g}^2\right\} = (\pi/2)A^2$ and (2) becomes $\eta^2 = (\pi/2)(\sigma_d^2/\max\{\sigma_{y_g}^2\}) = (1/A^2)W^{o^{\mathrm{T}}} R_{xx} W^o$

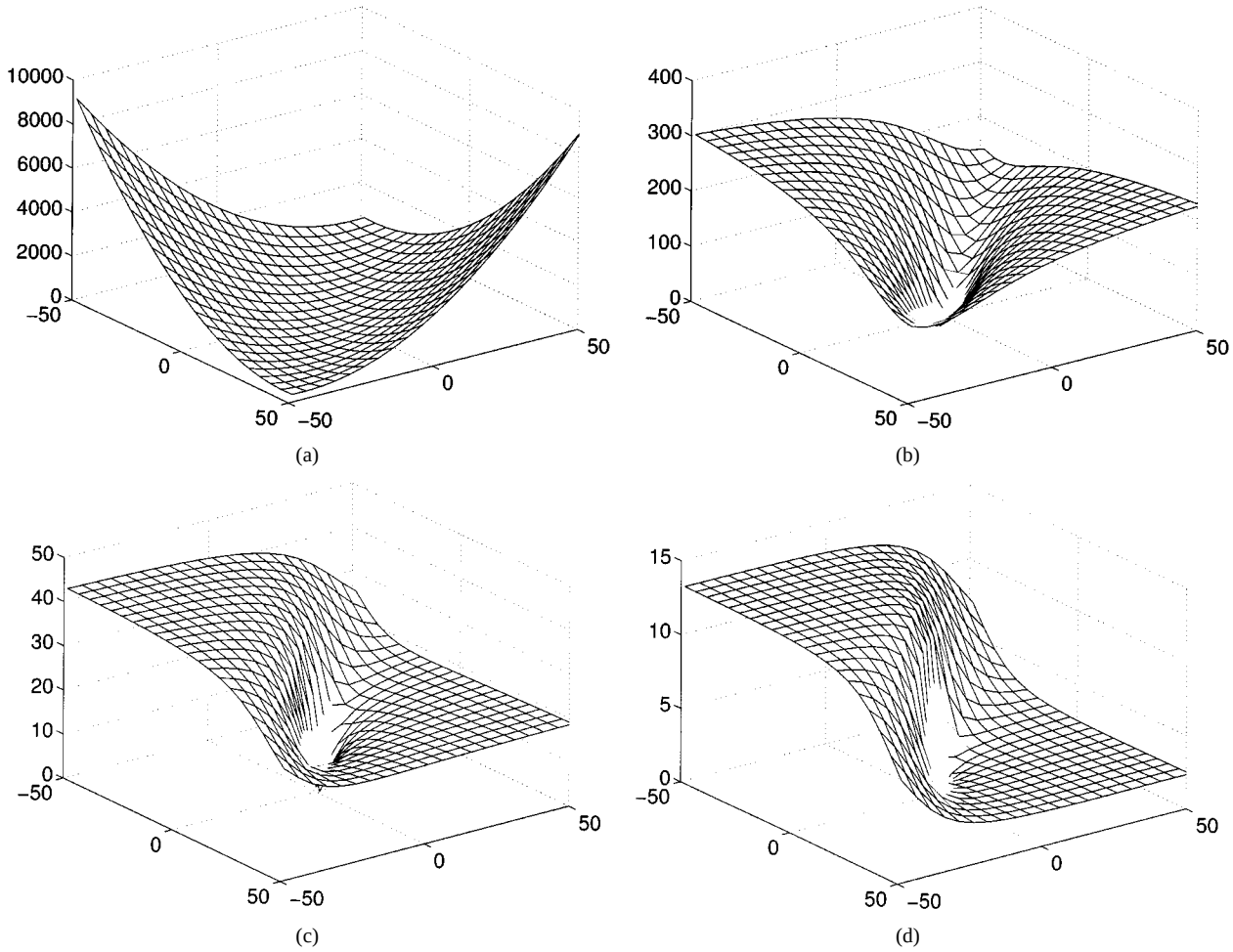


Fig. 3. Mean-square error performance surface for different degrees of nonlinearity. $W^o = [0.707\ 0.707]$; $\sigma_x^2 = 1$; $\sigma_z^2 = 10^{-6}$; eigenvalue spread of R_{xx} equal to 24. (a) MSE contour for $\eta^2 = 10^{-5}$. (b) $\eta^2 = 0.01$. (c) $\eta^2 = 0.1$. (d) $\eta^2 = 0.5$.

Squaring $e(n)$ obtained from Fig. 2 and taking the expected value yields

$$\begin{aligned} E\left\{e^2(n)\right\} = {} & W^{o^{\mathrm{T}}} E\left\{X(n)X^{\mathrm{T}}(n)\right\} W^o + E\left\{z^2(n)\right\} \\ & + 2W^{o^{\mathrm{T}}} E\left\{z(n)X(n)\right\} \\ & - 2W^{o^{\mathrm{T}}} E\left\{g\left[W^{\mathrm{T}}X(n)\right] X(n)\right\} \\ & - 2E\left\{g\left[W^{\mathrm{T}}X(n)\right] z(n)\right\} \\ & + E\left\{g^2\left[W^{\mathrm{T}}X(n)\right]\right\}. \end{aligned} \tag{4}$$

The first three and the fifth expectations in (4) are easily evaluated using the statistical properties of $x(n)$, $z(n)$, and $d(n)$. Thus, $E\{X(n)X^{\mathrm{T}}(n)\} = R_{xx}$, $E\{z^2(n)\} = \sigma_z^2$, $E\{z(n)X(n)\} = 0$, and $E\{g[W^{\mathrm{T}}X(n)]z(n)\} = 0$. The remaining terms are expectations of functions of zero-mean jointly Gaussian variables. The fourth expectation can be obtained from [17, (A19)] for $b_1 = 0$, $\sigma_q = \sigma$, $c = 1/\sigma$, and $\sigma_y^2 = W^{\mathrm{T}}R_{xx}W$. Thus

$$E\left\{g\left[W^{\mathrm{T}}X(n)\right] X(n)\right\} = \frac{1}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W + 1}} R_{xx}W. \tag{5}$$

The last expectation can be obtained from [18, (40)] for $\alpha_f = b_1(n) = 1$ and $b = W^{\mathrm{T}}R_{xx}W$ as

$$E\left\{g^2\left[W^{\mathrm{T}}X(n)\right]\right\} = \sigma^2 \arcsin\left(\frac{W^{\mathrm{T}}R_{xx}W}{W^{\mathrm{T}}R_{xx}W + \sigma^2}\right). \tag{6}$$

Combining the above results into (4) yields an analytical expression for the MSE surface

$$\begin{aligned} \xi = E\left\{e^2(n)\right\} = {} & W^{o^{\mathrm{T}}} R_{xx} W^o - \frac{2}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W + 1}} \\ & \times W^{o^{\mathrm{T}}} R_{xx} W + \sigma^2 \arcsin\left(\frac{W^{\mathrm{T}}R_{xx}W}{W^{\mathrm{T}}R_{xx}W + \sigma^2}\right) + \sigma_z^2. \end{aligned} \tag{7}$$

Equation (7) reduces to the MSE expression for the linear case as $\sigma^2 \to \infty$ [3]. Fig. 3 shows examples of the MSE surface for different degrees of nonlinearity η^2. Notice that the surface deforms as η^2 increases, but appears to remain unimodal. This important result will be demonstrated in the next subsection.

C. Stationary Points

R_{xx} is assumed positive definite, which is a reasonable assumption for most practical systems [3]. Differentiating (7) with respect to W, equating the result to zero, and denoting $\tilde{W}$ as the

finite values of W that satisfy the resulting equation, it can be easily shown that

$$\tilde{W}=\frac{\left(1+\frac{1}{\sigma^2}\tilde{W}^{\mathrm{T}}R_{xx}\tilde{W}\right)}{\frac{1}{\sigma^2}W^{o\mathrm{T}}R_{xx}\tilde{W}+\frac{\left(\frac{1}{\sigma^2}\tilde{W}^{\mathrm{T}}R_{xx}\tilde{W}+1\right)^{1/2}}{\left(\frac{2}{\sigma^2}\tilde{W}^{\mathrm{T}}R_{xx}\tilde{W}+1\right)^{1/2}}}W^o=cW^o. \quad (8)$$

Note that the multiplier c in (8) is a real scalar for any finite W^o and $\tilde{W}$. Thus, $\tilde{W}$ is a scaled version of W^o. This result is in agreement with the result derived in [19] for a single perceptron. Substituting cW^o for $\tilde{W}$ in (8) and using (2) yields

$$cW^o=\frac{c^2\eta^2+1}{c\eta^2+\frac{(c^2\eta^2+1)^{1/2}}{(2c^2\eta^2+1)^{1/2}}}W^o. \quad (9)$$

Equating the scalar multiples in both sides of (9) yields

$$c\sqrt{\frac{c^2\eta^2+1}{2c^2\eta^2+1}}=1 \quad (10)$$

which shows that c must be positive. Squaring (10) and solving for c yields the four solutions

$$c_{1,2,3,4}=\pm\sqrt{1-\frac{1}{2\eta^2}\pm\sqrt{\frac{1}{4\eta^4}+1}}. \quad (11)$$

It is easy to verify that the only solution satisfying $c\in\mathbb{R}^+$ is

$$c=\sqrt{1-\frac{1}{2\eta^2}+\sqrt{\frac{1}{4\eta^4}+1}}. \quad (12)$$

Equation (12) shows that

$$\tilde{W}=cW^o=\left[\sqrt{1-\frac{1}{2\eta^2}+\sqrt{\frac{1}{4\eta^4}+1}}\right]W^o \quad (13)$$

corresponds to the only finite point for which $\partial\xi/\partial W=0$. Appendix A presents a mathematical proof that the Hessian $\nabla^2[\xi(W)]$ is positive definite at $W=\tilde{W}$. Thus, (13) corresponds to a minimum of $\xi(W)$. Fig. 4 shows the multiplicative bias c for a large range of η^2.

Setting $W=\tilde{W}$ in (7) and using (13) yields an expression for the minimum MSE

$$\xi_{min}=\sigma_z^2+\left[1-\frac{2c}{\sqrt{c^2\eta^2+1}}+\frac{1}{\eta^2}\arcsin\left(\frac{c^2\eta^2}{c^2\eta^2+1}\right)\right]\times W^{o^{\mathrm{T}}}R_{xx}W^o. \quad (14)$$

Again, as $\eta^2\to 0$, (14) reduces to the linear case optimum solution $\xi_{\min}=\sigma_z^2$. Fig. 5 shows the excess MSE (the additional loss in cancellation level due to the nonlinearity) relative to the linear case ($\xi_{\min}-\sigma_z^2$) as a function of η^2 for $W^{o^{\mathrm{T}}}R_{xx}W^o=1$. Figs. 4 and 5 show the significant impact of the nonlinearity on the achievable cancellation level as compared with the bias of the optimum weight vector.

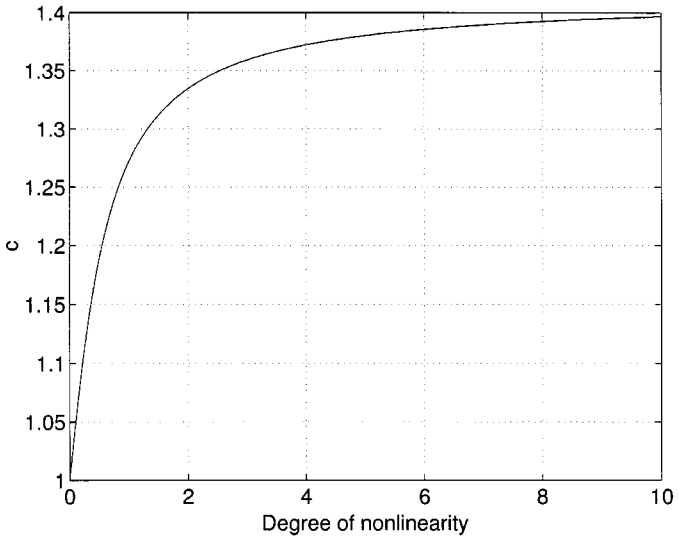


Fig. 4. Optimum weight vector multiplicative bias as a function of η^2.

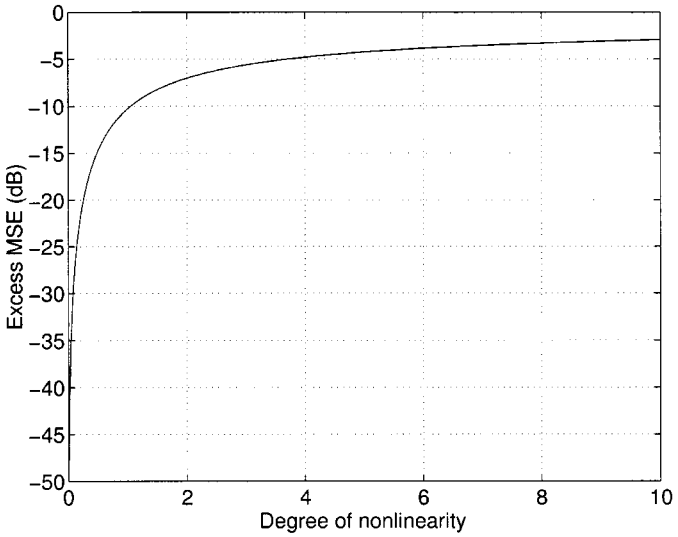


Fig. 5. Steady-state excess MSE relative to the linear case ($\xi_{\min}-\sigma_z^2$) as a function of η^2.

III. Analysis–LMS Algorithm Transient Behavior

This section analyzes the transient LMS algorithm behavior for Fig. 1. The weight vector is time-varying ($W(n)$) and is adjusted with the LMS algorithm. Thus, $W(n)=[w_0(n)\ w_1(n)\cdots w_{N-1}(n)]^{\mathrm{T}}$.

A. Mean Weight Behavior

The weight update equation for the LMS algorithm is given by

$$W(n+1)=W(n)+\mu e(n)X(n) \quad (15)$$

where μ is the adaptation step size, and

$$\begin{aligned}e(n)&=d(n)+z(n)-y_g(n)\\&=W^{o^{\mathrm{T}}}X(n)+z(n)-g\left[W^{\mathrm{T}}(n)X(n)\right].\end{aligned} \quad (16)$$

Using (16) in (15) yields

$$\begin{aligned}W(n+1)=W(n)+\mu\{&W^{oT}X(n)+z(n)\\&-g[W^{\mathrm{T}}(n)X(n)]\}X(n).\end{aligned} \quad (17)$$

The expected value of (17) is obtained in two steps. First, the expectation is taken conditioned on $W(n)$, leading to the recursion

$$\begin{aligned} E\{W(n+1)|W(n)\} &= W(n)+\mu E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}W^{o} \\ &\quad +\mu E\{z(n)X(n)|W(n)\} \\ &\quad -\mu E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\}. \end{aligned} \tag{18}$$

For sufficiently small μ, the first conditional expectation on the right-hand side of (18) can be approximated [20], [21] by

$$E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\} \approx E\left\{X(n)X^{\mathrm{T}}(n)\right\} = R_{xx}. \tag{19}$$

The second conditional expectation on the right-hand side of (18) is zero since $z(n)$ is statistically independent of $X(n)$. The third conditional expectation is [see (5)]

$$\begin{aligned} &E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\} \\ &\quad = \frac{1}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}R_{xx}W(n). \end{aligned} \tag{20}$$

Substituting (19) and (20) into (18) yields

$$\begin{aligned} &E\{W(n+1)|W(n)\} \\ &\quad = \left[I-\mu\frac{1}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}R_{xx}\right] \\ &\qquad \times W(n)+\mu R_{xx}W^{o}. \end{aligned} \tag{21}$$

Since the joint probability density function of the vector $W(n)$ is not known, the expected value of (21) can only be approximated. The following approximation is used:

$$\begin{aligned} &E\{W(n+1)\} \\ &\quad \approx \left[I-\mu\frac{1}{\sqrt{\frac{1}{\sigma^2}E\left\{W^{\mathrm{T}}(n)R_{xx}W(n)\right\}+1}}R_{xx}\right] \\ &\qquad \times E\{W(n)\}+\mu R_{xx}W^{o} \\ &\quad = \left[I-\mu\frac{1}{\sqrt{\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right]+1}}R_{xx}\right] \\ &\qquad \times E\{W(n)\}+\mu R_{xx}W^{o} \end{aligned} \tag{22}$$

where $W(n)$ and $W^{\mathrm{T}}(n)R_{xx}W(n)$ have been approximated by their expected values. $\mathrm{tr}[\cdot]$ stands for the trace of the matrix. Note that (22) reduces to the mean weight equation for the linear case as $\sigma^2 \to \infty$. An approximate recursive expression for $E\{W(n)W^{\mathrm{T}}(n)\}$ will be found in the next subsections.

B. Mean Square Error Behavior

Squaring (16) and taking the conditional expectation given $W(n)$ yields

$$\begin{aligned} E\left\{e^2(n)|W(n)\right\} &= W^{o^{\mathrm{T}}}E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}W^{o} \\ &\quad +2W^{o^{\mathrm{T}}}E\{z(n)X(n)|W(n)\} \\ &\quad -2W^{o^{\mathrm{T}}}E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\} \\ &\quad +E\left\{z^2(n)|W(n)\right\} \\ &\quad -2E\left\{z(n)g\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\} \\ &\quad +E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\}. \end{aligned} \tag{23}$$

The first expectation has been evaluated in (19). The second and the fifth expectations are equal to zero because $z(n)$ is zero mean and independent. The third expectation has been evaluated in (20). The fourth one is equal to σ_z^2. The last term follows directly from (6) by replacing W with $W(n)$. Thus

$$\begin{aligned} &E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\} \\ &\quad = \sigma^2\arcsin\left(\frac{W^{\mathrm{T}}(n)R_{xx}W(n)}{W^{\mathrm{T}}(n)R_{xx}W(n)+\sigma^2}\right). \end{aligned} \tag{24}$$

Substituting (19), (20), and (24) in (23) and rearranging the terms yields

$$\begin{aligned} E\left\{e^2(n)|W(n)\right\} = &-\frac{2}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}} \\ &\times W^{o^{\mathrm{T}}}R_{xx}W(n) \\ &+\sigma^2\arcsin\left(\frac{W^{\mathrm{T}}(n)R_{xx}W(n)}{W^{\mathrm{T}}(n)R_{xx}W(n)+\sigma^2}\right) \\ &+W^{o^{\mathrm{T}}}R_{xx}W^{o}+\sigma_z^2. \end{aligned} \tag{25}$$

The evaluation of the expected value of (25) over $W(n)$ is not a simple task because the density function of the weight vector is not known. It can be approximated by

$$\begin{aligned} \xi(n) = E\left\{e^2(n)\right\} &\approx -\frac{2}{\sqrt{\frac{1}{\sigma^2}E\left\{W^{\mathrm{T}}(n)R_{xx}W(n)\right\}+1}} \\ &\quad \times W^{o^{\mathrm{T}}}R_{xx}E\{W(n)\} \\ &\quad +\sigma^2\arcsin\left(\frac{E\left\{W^{\mathrm{T}}(n)R_{xx}W(n)\right\}}{E\left\{W^{\mathrm{T}}(n)R_{xx}W(n)\right\}+\sigma^2}\right) \\ &\quad +W^{o^{\mathrm{T}}}R_{xx}W^{o}+\sigma_z^2 \\ &= -\frac{2}{\sqrt{\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right]+1}} \\ &\quad \times W^{o^{\mathrm{T}}}R_{xx}E\{W(n)\} \\ &\quad +\sigma^2\arcsin\left(\frac{\mathrm{tr}\left[R_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right]}{\mathrm{tr}\left[R_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right]+\sigma^2}\right) \\ &\quad +W^{o^{\mathrm{T}}}R_{xx}W^{o}+\sigma_z^2. \end{aligned} \tag{26}$$

As $\sigma^2 \to \infty$ (linear case), (26) converges to the MSE expression for the linear case [3]:

$$\xi(n) = W^{o^{\mathrm{T}}} R_{xx} W^o + \sigma_z^2 - 2W^{o^{\mathrm{T}}} R_{xx} E\{W(n)\} + \mathrm{tr}\left[R_{xx} E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right]. \quad (27)$$

C. Weight Correlation Matrix

Evaluation of (22) and (26) requires $E\{W(n)W^{\mathrm{T}}(n)\}$. A recursion for the conditional weight correlation matrix is derived in Appendix B as (28), shown at the bottom of the page. The expectation of (28) over $W(n)$ represents a formidable mathematical task. Approximate expressions can be obtained using numerous different approaches. The following approximations preserve information about the first and second moments of the adaptive weights in the dominant terms of (28) while keeping the mathematical problem tractable. These approximations for a stochastic model of the adaptive algorithm behavior are supported (even for reasonably large values of μ) by the simulation results presented in Section VI

$$\begin{aligned}
&W(n)W^{\mathrm{T}}(n) = E\left\{W(n)W^{\mathrm{T}}(n)\right\} + \text{terms in } \mu^2 \\
&W^{\mathrm{T}}(n)R_{xx}W(n) = \mathrm{tr}\left[R_{xx}W(n)W^{\mathrm{T}}(n)\right] \\
&\quad = \mathrm{tr}\left[R_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}\right] + \text{terms in } \mu^2 \\
&W(n) = E\left\{W(n)\right\} + \text{terms in } \mu \\
&W^{\mathrm{T}}(n)R_{xx}W^oR_{xx}W(n)W^{\mathrm{T}}(n)R_{xx} \\
&\quad = E\left\{W^{\mathrm{T}}(n)\right\}R_{xx}W^oR_{xx}E\left\{W(n)W^{\mathrm{T}}(n)\right\}R_{xx} \\
&\qquad + \text{terms in } \mu^2. \quad (29)
\end{aligned}$$

Neglecting terms in μ^k for $k \geq 1$ in (29) and letting $K(n) = E\left\{W(n)W^{\mathrm{T}}(n)\right\}$ yields (30), shown at the bottom of the next page. Equations (22), (26), and (30) form the analytical model for studying the statistical behavior of the LMS adaptive algorithm in the system of Fig. 1.

IV. Simplified Model-White Signals and Slow Adaptation

The analytical model derived above can be specialized for a white input signal by setting $R_{xx} = \sigma_x^2 I$. However, further simplifications are possible for white inputs and very small μ. The importance of such a simplified model is twofold: i) The white input case with small step-size represents an important share of practical applications, mainly in system identification, and ii) the analytical model reduces to scalar recursions. These are easy to handle and lead to interesting insights into the algorithm behavior. The white-input-small-μ algorithm behavior can serve as a baseline for other cases. Larger step sizes speed up convergence but with an increase in steady-state cancellation level. Signal correlation can slow-down convergence.

Consider $x(n)$ white and μ sufficiently small so that the effects of weight fluctuations can be neglected in (22). Thus, $R_{xx} = \sigma_x^2 I$ and (22) can be further approximated by

$$E\{W(n+1)\} = \left[1 - \mu \frac{\sigma_x^2}{\sqrt{\frac{\sigma_x^2}{\sigma^2}E\{W^{\mathrm{T}}(n)\}E\{W(n)\} + 1}}\right] \times E\{W(n)\} + \mu\sigma_x^2 W^o. \quad (31)$$

Consider the initialization weight vector $W(0) = 0$. From (31), $E\{W(1)\} = \mu\sigma_x^2 W^o$ in this case. Thus, $E\{W(n)\}$ is collinear with W^o at the first iteration. Moreover, $E\{W(n)\}$ will always follow the direction of W^o since the term multiplying $E\{W(n)\}$ in (31) is a scalar. Hence, (31) can be written as

$$E\{W(n)\} = k(n)W^o \text{ for all } n \quad (32)$$

$$\begin{aligned}
E\left\{W(n+1)W^{\mathrm{T}}(n+1)|W(n)\right\} &= W(n)W^{\mathrm{T}}(n) + \mu W(n)W^{o^{\mathrm{T}}}R_{xx} + \mu R_{xx}W^oW^{\mathrm{T}}(n) \\
&- \frac{\mu}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}R_{xx}W(n)W^{\mathrm{T}}(n) - \frac{\mu}{\sqrt{\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}W(n)W^{\mathrm{T}}(n)R_{xx} \\
&+ \mu^2\sigma_z^2R_{xx} + 2\mu^2R_{xx}W^oW^{o^{\mathrm{T}}}R_{xx} + \mu^2W^{o^{\mathrm{T}}}R_{xx}W^oR_{xx} \\
&+ \frac{2\mu^2W^{\mathrm{T}}(n)R_{xx}W^oR_{xx}W(n)W^{\mathrm{T}}(n)R_{xx}}{\sigma^2\left[\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right]^{3/2}} \\
&- \frac{2\mu^2\left[W^{\mathrm{T}}(n)R_{xx}W^oR_{xx} + R_{xx}W(n)W^{o^{\mathrm{T}}}R_{xx} + R_{xx}W^oW^{\mathrm{T}}(n)R_{xx}\right]}{\left[\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right]^{1/2}} \\
&+ \mu^2\sigma^2 \arcsin\left(\frac{W^{\mathrm{T}}(n)R_{xx}W(n)}{W^{\mathrm{T}}(n)R_{xx}W(n)+\sigma^2}\right)R_{xx} \\
&+ \frac{2\mu^2R_{xx}W(n)W^{\mathrm{T}}(n)R_{xx}}{\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)\sqrt{\frac{2}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}. \quad (28)
\end{aligned}$$

and the recursive equation becomes scalar. Substituting (32) in (31) and using (2) leads to

$$k(n+1)=\left[1-\mu\frac{\sigma_x^2}{\sqrt{\eta^2k^2(n)+1}}\right]k(n)+\mu\sigma_x^2. \tag{33}$$

Applying the same assumptions to (26) and using (2), $W(0)=0$ and $E\{W(n)\}=k(n)W^o$ leads to

$$\xi(n)=\sigma_x^2W^{o^{\mathrm{T}}}W^o\left[1-\frac{2}{\sqrt{\eta^2k^2(n)+1}}k(n)+\frac{1}{\eta^2}\arcsin\left(\frac{\eta^2k^2(n)}{\eta^2k^2(n)+1}\right)\right]+\sigma_z^2. \tag{34}$$

Equations(32)–(34) determine $E\{W(n)\}$ and $E\{e^2(n)\}$ for all n when the adaptive filter is initialized at $W(0)=0$, the input signal $x(n)$ is white, and the step size is small.

V. Steady-State Algorithm Behavior

This section studies the limiting behavior of the converged LMS algorithm. The determination of the steady-state algorithm behavior from (22), (26), and (30) requires numerical methods. However, using the assumption of very small weight fluctuations (compared with their mean values), very simple analytical expressions that are useful for evaluation and design purposes can be determined. It is then assumed that $\lim_{n\to\infty}W(n)=\lim_{n\to\infty}E\{W(n)\}=W_\infty$ for the steady-state analysis.

A. Mean Weight Steady-State Behavior

Assume algorithm convergence as $n\to\infty$. Replacing $W(n)$ with W_∞ in (22) yields

$$W_\infty=W^o\sqrt{\frac{1}{\sigma^2}W_\infty^{\mathrm{T}}R_{xx}W_\infty+1}. \tag{35}$$

Since R_{xx} positive semi-definite [3], $W_\infty^{\mathrm{T}}R_{xx}W_\infty\geq 0$.[3] Thus, it is clear from (35) that $W_\infty=aW^o$, where $a\in\mathbb{R}^+$. Substituting aW^o for W_∞ in (35), solving for $a\in\mathbb{R}^+$, and using (2) yields

$$W_\infty=\lim_{n\to\infty}E\{W(n)\}=\frac{1}{\sqrt{1-\eta^2}}W^o \tag{36}$$

which shows that $E\{W(n)\}$ converges to a scaled version of the unknown system's response for Gaussian input signals. The identification error caused by the nonlinearity increases as $\eta^2\to 1$. This steady-state error cannot be reduced by reducing the adaptation step size. As $\eta^2\to 1$, $E\{W(n)\}$ generated by the LMS algorithm grows without limit. Equation (36) has no stationary points for $\eta^2>1$. This multiplicative bias occurs because the instantaneous approximation of the MSE used to derive the LMS algorithm update equation does not consider the nonlinearity effect.

Notice from (2) that $\eta^2=1$ establishes a power threshold[4] ($\max\{\sigma_{y_g}^2\}=(\pi/2)\sigma_d^2$). Above this threshold, the adaptive branch (including the nonlinearity) cannot provide sufficient signal power to cancel the power in the desired signal $d(n)$ (i.e., the adaptive algorithm is not able to increase the filter gain sufficiently to overcome the nonlinear saturation). Hence, the adaptive filter gain increases without bound.

In addition, notice that as $\sigma^2\to\infty$ (toward the linear case), $\eta^2\to 0$, and (36) reduces to the steady-state mean converged weight vector for the linear case.

B. Steady-State MSE

An approximate expression for the steady-state MSE behavior is determined by replacing $W(n)$ with the steady-state

[3]Positive definiteness of R_{xx} is required for the singular case of $\sigma=0$

[4]It can be readily shown that the threshold $\max\{\sigma_{y_g}^2\}=(\pi/2)\sigma_d^2$ is valid even if $g(y)$ in (1) is scaled by a nonzero real constant.

$$\begin{aligned}K(n+1)=K(n)&+\mu E\{W(n)\}W^{o^{\mathrm{T}}}R_{xx}+\mu R_{xx}W^oE\left\{W^{\mathrm{T}}(n)\right\}\\&-\frac{\mu}{\left(\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)^{1/2}}R_{xx}K(n)-\frac{\mu}{\left(\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)^{1/2}}K(n)R_{xx}\\&+\mu^2\sigma_z^2R_{xx}+2\mu^2R_{xx}W^oW^{o^{\mathrm{T}}}R_{xx}+\mu^2W^{o^{\mathrm{T}}}R_{xx}W^oR_{xx}\\&+\frac{2\mu^2E\left\{W^{\mathrm{T}}(n)\right\}R_{xx}W^oR_{xx}K(n)R_{xx}}{\sigma^2\left(\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)^{3/2}}\\&-\frac{2\mu^2\left[E\left\{W^{\mathrm{T}}(n)\right\}R_{xx}W^oR_{xx}+R_{xx}E\{W(n)\}W^{o^{\mathrm{T}}}R_{xx}+R_{xx}W^oE\left\{W^{\mathrm{T}}(n)\right\}R_{xx}\right]}{\left(\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)^{1/2}}\\&+\mu^2\sigma^2\arcsin\left(\frac{\mathrm{tr}\left[R_{xx}K(n)\right]}{\mathrm{tr}\left[R_{xx}K(n)\right]+\sigma^2}\right)R_{xx}\\&+\frac{2\mu^2R_{xx}K(n)R_{xx}}{\left(\frac{1}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)\left(\frac{2}{\sigma^2}\mathrm{tr}\left[R_{xx}K(n)\right]+1\right)^{1/2}}.\end{aligned} \tag{30}$$

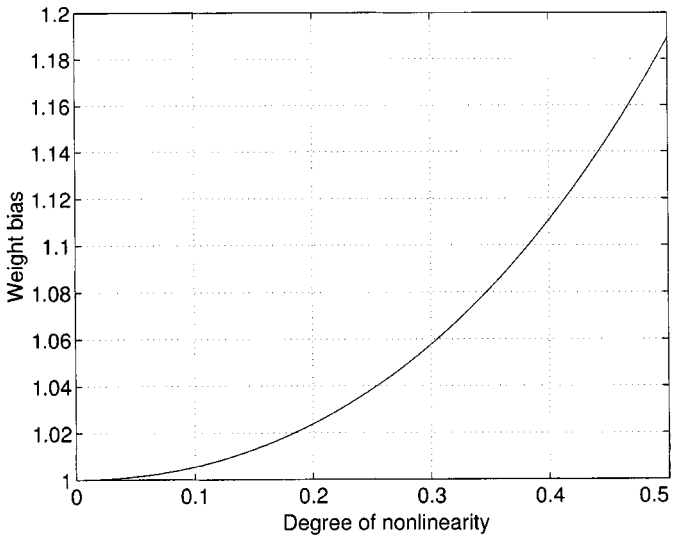


Fig. 6. Weight bias κ ($W_\infty = \kappa \tilde{W}$) caused by the LMS algorithm as a function of η^2.

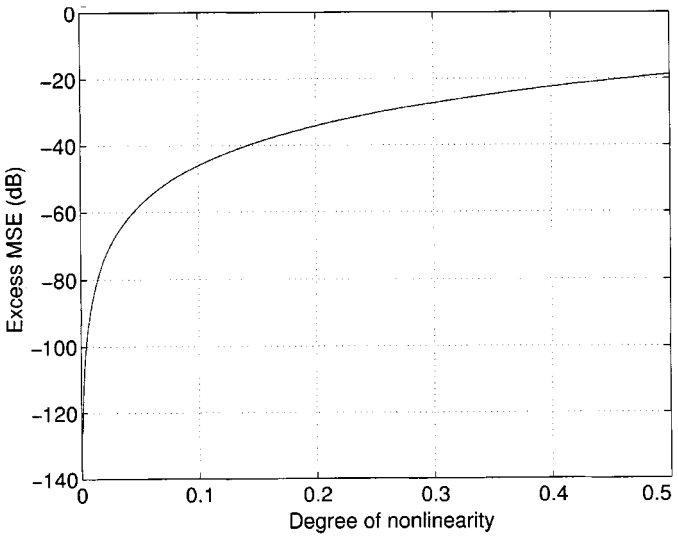


Fig. 7. Steady-state excess MSE for the LMS algorithm as a function of η^2.

mean weight vector expression (36) in (26). After some simple algebraic manipulations

$$\lim_{n\to\infty} \xi(n) = \lim_{n\to\infty} E\left\{e^2(n)\right\} = W^{o^{\mathrm{T}}} R_{xx} W^o \left(\frac{1}{\eta^2}\arcsin\left(\eta^2\right) - 1\right) + \sigma_z^2. \quad (37)$$

The first term in (37) is the effect of the nonlinearity on the LMS algorithm steady-state MSE for small μ. As $\eta^2 \to 0$, $\lim_{n\to\infty} \xi(n) = \sigma_z^2$ (the minimum MSE for the linear LMS algorithm with slow adaptation). Equations (36) and (37) show how nonlinearity affects the linear LMS algorithm steady-state behavior. These expressions must also be compared with the results obtained for the MSE performance surface to determine the weight vector bias and the excess MSE. These results are significant because the nonlinearity $g(\cdot)$ is usually inherent to the system.

The LMS multiplicative weight bias κ is obtained from (13) and (36) as

$$W_\infty = \left[\frac{2\eta^2}{(1-\eta^2)\sqrt{1+4\eta^4} - 2\eta^4 + 3\eta^2 - 1}\right]^{1/2} \tilde{W} = \kappa \tilde{W}. \quad (38)$$

As $\eta^2 \to 0$ (linear case), $\kappa \to 1$. When $\eta^2 \to 1$, κ and W_∞ grow without bound. Fig. 6 shows the fast increase in the converged multiplicative weight bias as a function of η^2.

The steady-state excess MSE (EMSE) of the LMS algorithm, relative to (14), is obtained from (14) and (37) as

$$\mathrm{EMSE} = \left\{\frac{1}{\eta^2}\left[\arcsin(\eta^2) - \arcsin\left(\frac{c^2\eta^2}{c^2\eta^2+1}\right)\right] + \frac{2c}{\sqrt{c^2\eta^2+1}} - 2\right\} W^{o^{\mathrm{T}}} R_{xx} W^o \quad (39)$$

with c defined in (12). Fig. 7 shows the steady-state EMSE for $0 \le \eta^2 < 0.5$ and normalized signal power ($W^{o^{\mathrm{T}}} R_{xx} W^o = 1$). The impact of the nonlinearity must be compared with the linear case. This is because many practical systems use the LMS algorithm presuming the system linear. Such is the case in active noise control systems, for example. The nonlinear effect on the MSE surface adds to the minimization of the MSE using a stochastic gradient algorithm. The total deviation from the linear case combines the nonlinear effect on the MSE surface and the EMSE (39) resulting from using a stochastic gradient algorithm. $\xi(n)$ remains bounded as $\eta^2 \to 1$, even though the adaptive filter weights diverge. This behavior is due to the nature of the nonlinear saturation. It is easily shown that $g(y) = \sqrt{(\pi/2)}\sigma \mathrm{erf}(y/\sigma\sqrt{2})$. Thus, $\lim_{y\to\infty} g(y) = \sqrt{(\pi/2)}\sigma$, which leads to the bounded nonlinearity output and bounded MSE.

VI. Simulation Results

This section presents some simulation examples in support of the assumptions used to derive the theoretical models. Some representative plots have been selected from a large set of results.

1) Example 1: Consider $W^o = [0.4130\ 0.4627\ 0.4803\ 0.4627\ 0.4130]^{\mathrm{T}}$, $W^{o^{\mathrm{T}}} W^o = 1$, $x(n)$ white with $\sigma_x^2 = 1$ and measurement noise $z(n)$ with $\sigma_z^2 = 10^{-6}$. Simulations are presented for three step sizes (normalized with respect to the linear LMS stability limit $\mu_{\max} = 2/(\mathrm{tr}[R_{xx}])$). 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, $\eta^2 = 0.0005$, 0.05, 0.3, and 0.5 have been selected to illustrate the model accuracy for small, moderate, and large degrees of nonlinearity. Different values of η^2 are used for weight and MSE behaviors to avoid superimposed curves in single plots.

Fig. 8(a)–(c) compares the simulated mean weight behavior with the analytical predictions using (22) and (30). Each plot presents the results for $\eta^2 = 0.0005$, 0.3, and 0.5 and a single μ. 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, predicted by (36), is very accurate, even for the large μ in Fig. 8(a). The predicted steady-state values for $w_3(n)$

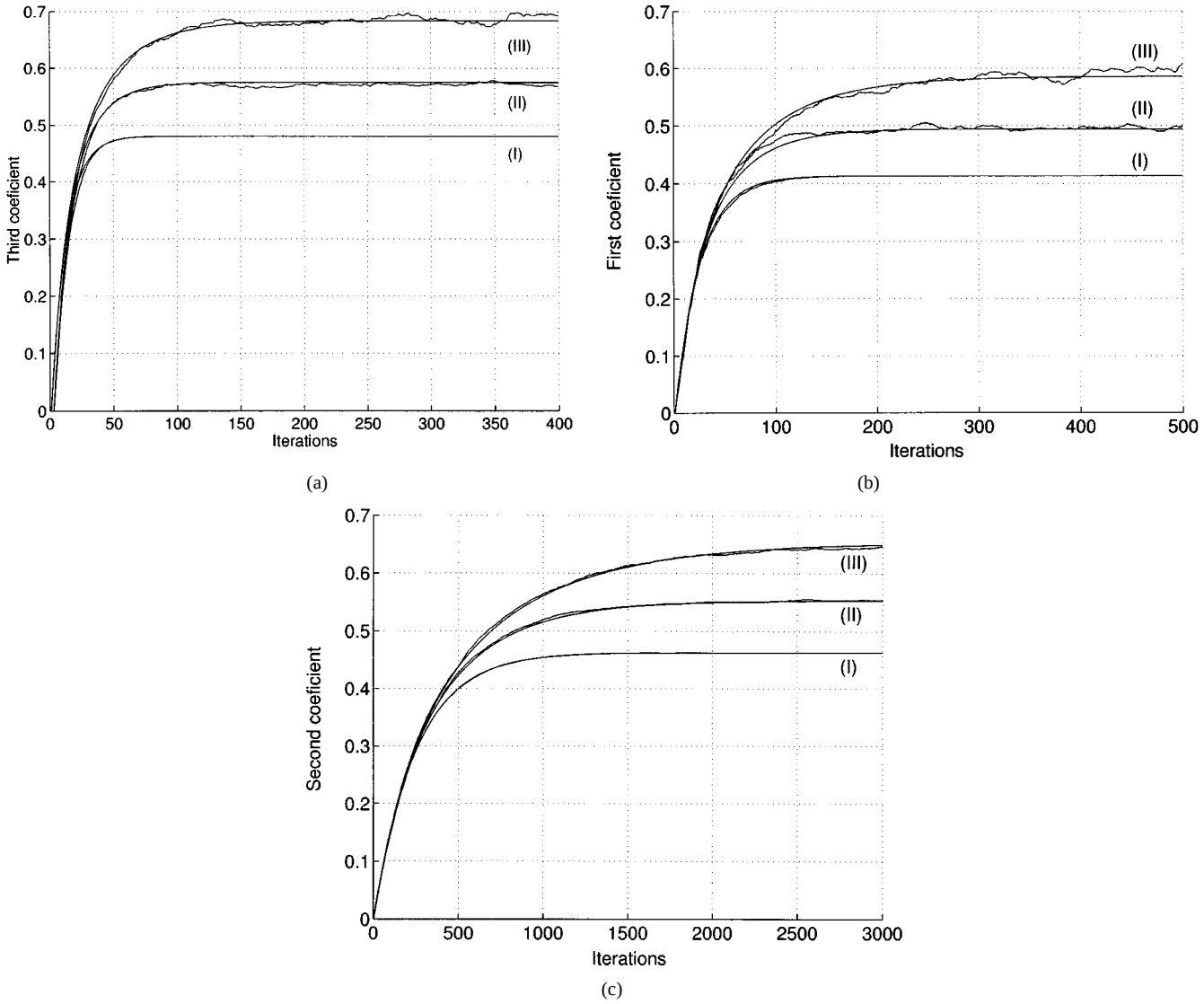


Fig. 8. Example 1: $E\{W(n)\}$ for $\eta^2 = 0.0005$ [curve (I)], 0.3 [curve (II)] and 0.5 [curve (III)]. Plots (a), (b), and (c) for different values of μ. Simulation—ragged curves. Theory—smooth curves. (a) $E\{w_3(n)\}$ for $\mu_1 = (\mu_{\max}/5) = 0.08$. (b) $E\{w_1(n)\}$ for $\mu_2 = (\mu_{\max}/10) = 0.04$. (c) $E\{w_2(n)\}$ for $\mu_3 = (\mu_{\max}/100) = 0.004$.

by (36) are 0.480, 0.574, and 0.679. Note that the weight fluctuations increase with η^2. This behavior is probably due to the saturation that clips the adaptive filter output signal for larger η^2. This clipping results in a larger error signal and to a larger weight update at each iteration.

Fig. 9(a), (c), and (e) show the simulated MSE and the theoretical predictions using (26) and (30). Each figure shows three curves, corresponding to $\eta^2 = 0.0005$, 0.05, and 0.5. Different plots are shown for different step sizes. Fig. 9(a) was obtained by averaging 1000 runs. Five hundred runs were averaged to obtain Fig. 9(c) and 9(e). The analytical model matches the simulations very well in all cases, even for the relatively large $\mu = \mu_1$. The steady-state MSE values [which were predicted by (37)] are -59.8 dB, -33.8 dB, and -13.3 dB. However, these values are clearly accurate only for small step sizes. Fig. 9 shows that the predicted steady-state values for the simplified model are closer to the simulation as μ decreases.

Fig. 9(a), (d), and (f) verify the accuracy of approximation (19) for different μ and η^2. Three vector components have been chosen at random to conserve space. All other components show similar behavior. The lines connecting the points are used for clarity only. Fig. 10(a) and (b) verify the accuracy of (32)–(34) for white inputs and small μ. Fig. 10(a) and (b) use the same signals and parameters as in Figs. 8(c) and 9(e), respectively. Figs. 8 and 9 show that increasing η^2 is more significant to the level of cancellation (steady-state MSE) than to the converged weight vector. The weight vector behavior for $\eta^2 = 0.05$ is very close to the behavior for $\eta^2 = 0.0005$ and, hence, is not shown. On the other hand, the steady-state MSE varies nearly by 30 dB as η^2 increases from 0.0005 to 0.05. This is mainly due to the distortion of the MSE surface, as demonstrated in Section II-B.

2) Example 2: This example repeats Example 1 for a correlated input signal. Thus, all the parameters, vectors, dimensions, and signal characteristics are the same as in Example 1, unless otherwise stated. The input signal $x(n)$ is a unit-variance autoregressive process obtained from a white Gaussian process so that the input vector $X(n)$ has an autocorrelation matrix R_{xx} with eigenvalue spread $\chi = 10.9$ [16].

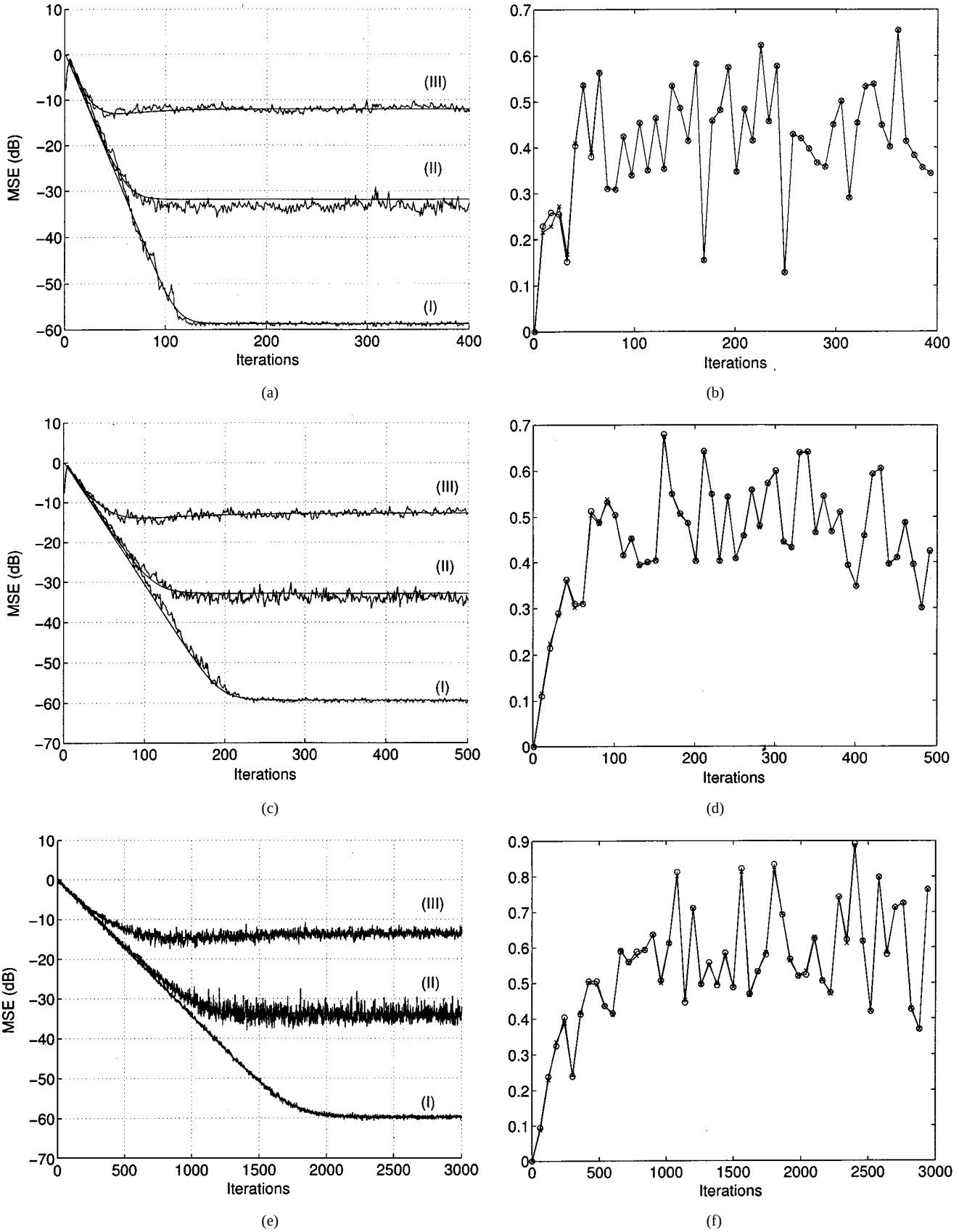
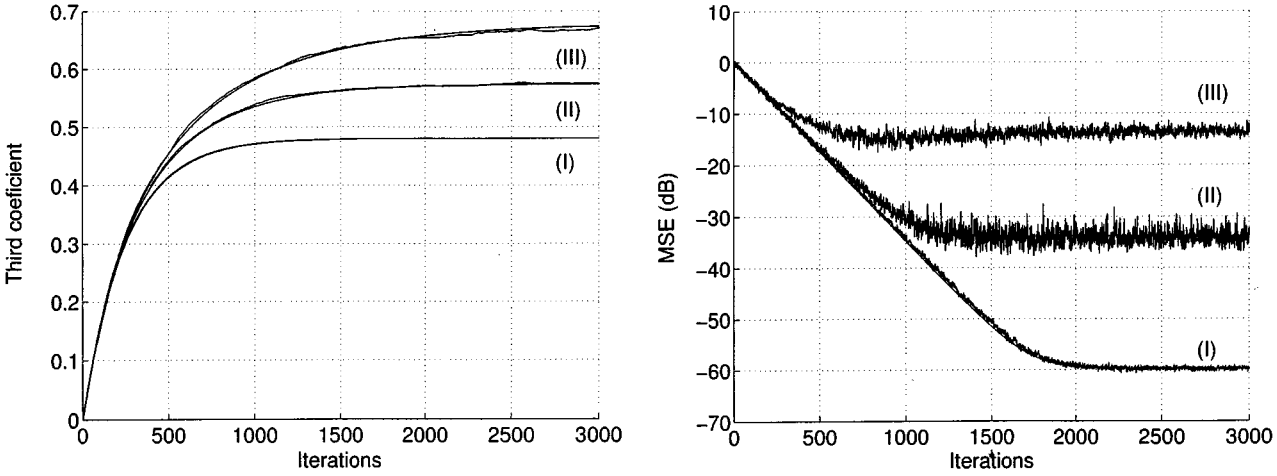
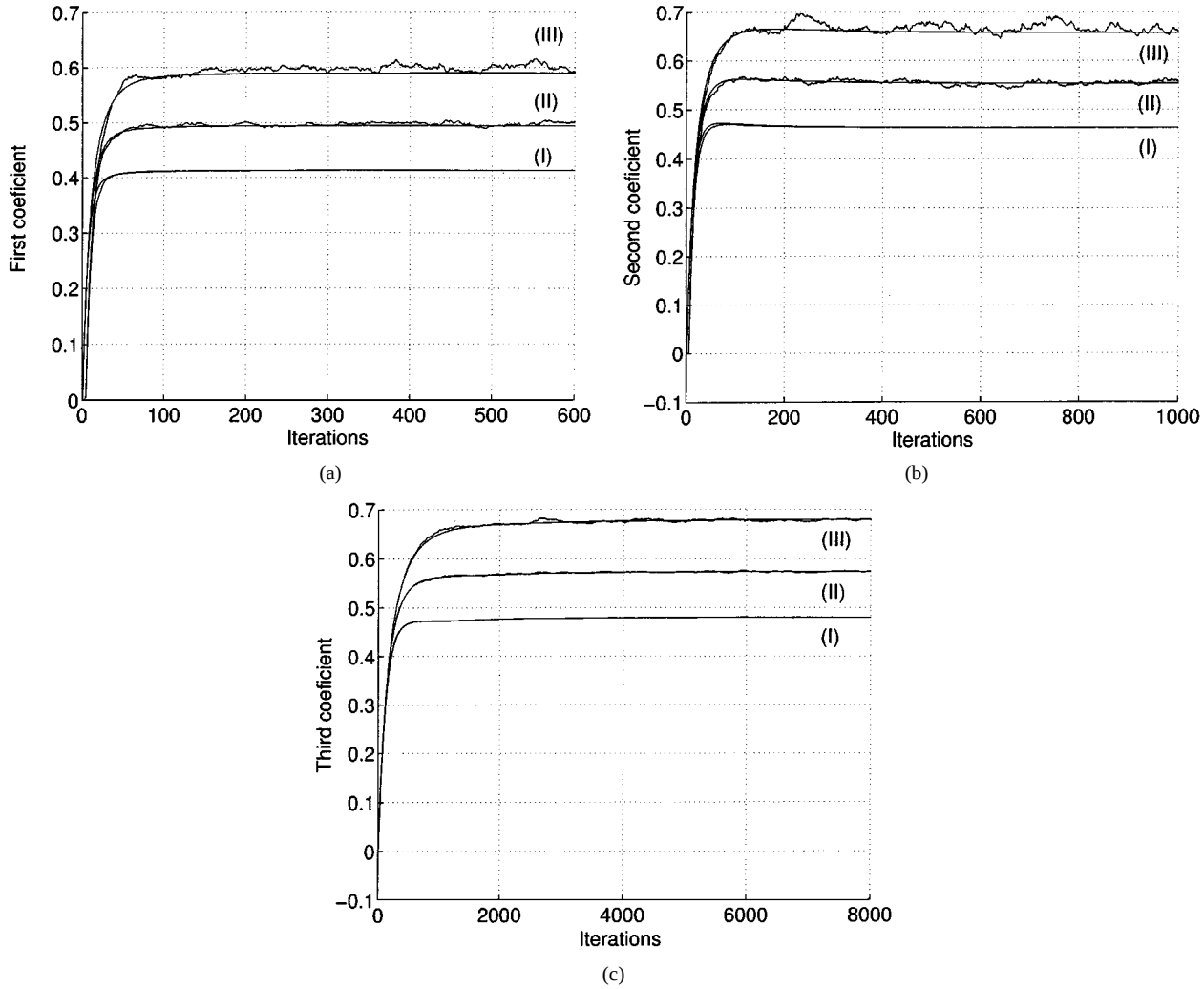


Fig. 9. Example 1: Left column: MSE for $\eta^2 = 0.0005$ [curve (I)], 0.05 [curve (II)] and 0.5 [curve (III)]. Simulation-ragged curves. Theory-smooth curves. Right column: Verification of (19) for $\eta^2 = 0.0005$, 0.05, and 0.5 (lines joining point only for clarity). $(\circ)E\{X(n)X^{\mathrm{T}}(n)W(n)\}$; $(\times)E\{X(n)X^{\mathrm{T}}(n)\}E\{W(n)\}$. (a) MSE for $\mu_1 = (\mu_{\max}/5) = 0.08$. (b) $\eta^2 = 0.0005$; $\mu_1 = (\mu_{\max}/5) = 0.08$; component 1. (c) MSE for $\mu_2 = (\mu_{\max}/10) = 0.04$. (d) $\eta^2 = 0.05$; $\mu_2 = (\mu_{\max}/10) = 0.04$; component 2. (e) MSE for $\mu_3 = (\mu_{\max}/100) = 0.004$, (f) $\eta^2 = 0.5$; $\mu_3 = (\mu_{\max}/100) = 0.004$; component 3.



(a) (b)

Fig. 10. Example 1: Verification of the simplified model. (a) $E\{w_3(n)\}$. Theory obtained from (33). (b) MSE with theory obtained from (33) and (34). All parameters and data identical to Figs. 8(c) and 9(e). (a) $E\{w_3(n)\}$ for $\mu = (\mu_{\max}/100) = 0.004$. (b) MSE for $\mu = (\mu_{\max}/100) = 0.004$.



(a) (b) (c)

Fig. 11. Example 2: $E\{W(n)\}$ for $\eta^2 = 0.0005$ [curve (I)], 0.3 [curve (II)] and 0.5 [curve (III)]. Plots (a), (b), and (c) for different values of μ. Simulation-ragged curves. Theory-smooth curves. (a) $E\{w_1(n)\}$ for $\mu_1 = (\mu_{\max}/5) = 0.08$. (b) $E\{w_2(n)\}$ for $\mu_2 = (\mu_{\max}/10) = 0.04$. (c) $E\{w_3(n)\}$ for $\mu_3 = (\mu_{\max}/100) = 0.004$.

Figs. 11 and 12 verify the analytical model using recursions (22), (26), and (30). The model accuracy is preserved for correlated input signals. The same conclusions are also valid as from Example 1.

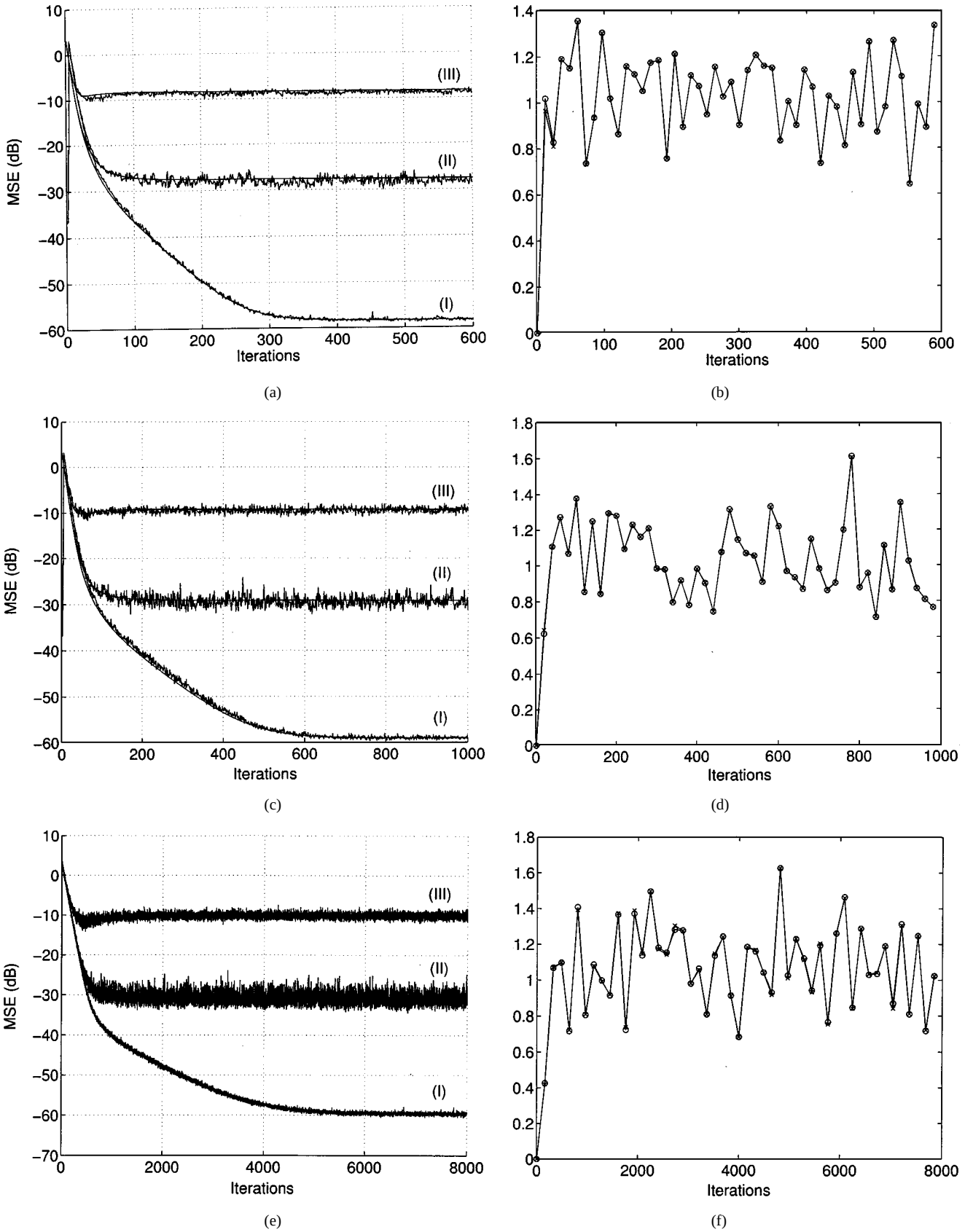


Fig. 12. Example 2: Left column: MSE for $\eta^2 = 0.0005$ [curve (I)], 0.05 [curve (II)] and 0.5 [curve (III)]. Simulation-ragged curves. Theory-smooth curves. Right column: Verification of (19) for $\eta^2 = 0.0005$, 0.05, and 0.5 (lines joining point only for clarity). $(\circ)E\{X(n)X^{\mathrm{T}}(n)W(n)\}$; $(\times)E\{X(n)X^{\mathrm{T}}(n)\}E\{W(n)\}$. (a) MSE for $\mu_1 = (\mu_{\max}/5) = 0.08$. (b) $\eta^2 = 0.0005$; $\mu_1 = (\mu_{\max}/5) = 0.08$; component 2. (c) MSE for $\mu_2 = (\mu_{\max}/10) = 0.04$. (d) $\eta^2 = 0.05$; $\mu_2 = (\mu_{\max}/10) = 0.04$; component 3. (e) MSE for $\mu_3 = (\mu_{\max}/100) = 0.004$, (f) $\eta^2 = 0.5$; $\mu_3 = (\mu_{\max}/100) = 0.004$; component 1.

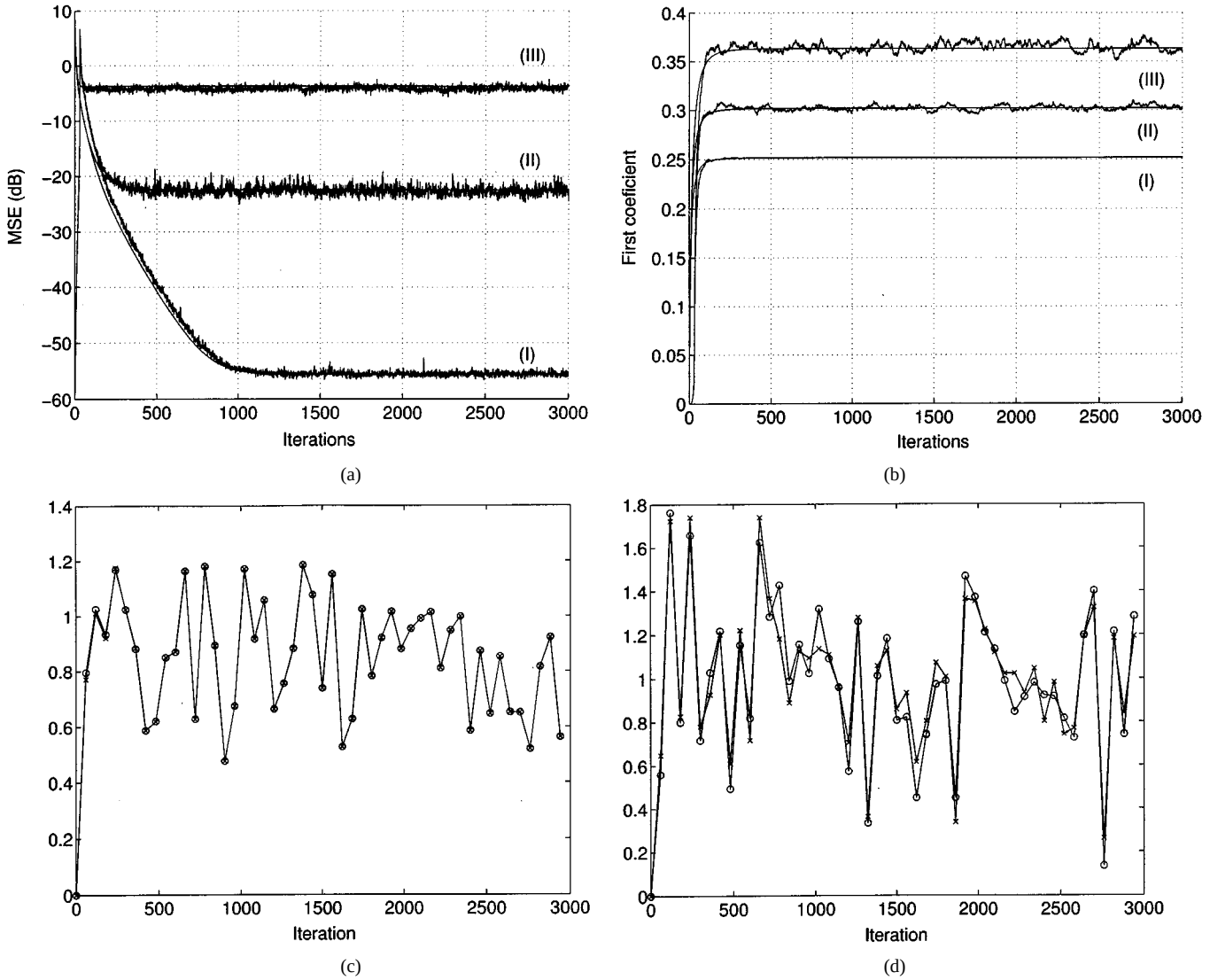


Fig. 13. Example 3: Algorithm behavior for highly correlated inputs and large step sizes. $\chi = 32.22$. Simulation-ragged curves. Theory-smooth curves. (a) MSE behavior. (b) Mean weight behavior. (c) and (d) Verification of (19). $(\circ)E\{X(n)X^{\mathrm{T}}(n)W(n)\}$; $(\times)E\{X(n)X^{\mathrm{T}}(n)\}E\{W(n)\}$. (a) MSE for $\mu = (\mu_{\max}/2) = 0.0333$ and $\eta^2 = 0.0005$ [curve (I)], 0.05 [curve (II)], and 0.5 [curve (III)]. (b) $E\{w_1(n)\}$ for $\mu = (\mu_{\max}/2) = 0.0333$ and $\eta^2 = 0.0005$ [curve (I)], 0.03 [curve (II)], and 0.5 [curve (III)]. (c) Verification of (19) for $\eta^2 = 0.0005$; $\mu = (\mu_{\max}/2) = 0.0333$ component 1. (d) Verification of (19) for $\eta^2 = 0.5$; $\mu = (\mu_{\max}/2) = 0.0333$ component 30.

3) Example 3: The last example considers a longer impulse response (30 taps going from $w_1^o = 0.252$ to $w_{30}^o = 0.103$ in steps of -0.005, unit norm) and a highly correlated input signal. The remaining parameters are unchanged unless explicitly stated. The input signal is a unit-variance autoregressive process with an eigenvalue spread of 32.22 [16]. The step size was large ($\mu = \mu_{\max}/2 = 0.0333$) in order to test the model in a very demanding situation.

Fig. 13 shows these results. Fig. 13(a) shows the MSE behavior. There is a small mismatch during the transient phase of adaptation and should be expected for such large μ. Otherwise, the model predicts the algorithm behavior very well. Fig. 13(b) shows the mean weight behavior. The large weight fluctuations for large η^2 are evident again. Fig. 13(c) and (d) verify (19). Fig. 13(c) is for the first vector component (largest in magnitude). Fig. 13(d) is for the 30th vector component (smallest in magnitude). The behavior of $w_{30}(n)$ is much more dependent on the input signal fluctuations, especially for such large step sizes. Thus, as seen in Fig. 13(d), this mismatch should be expected. This example represents a very extreme case. Note also that the analytical model is quite robust to large deviations from the assumptions used to derive the theory.

VII. Conclusion

This paper has presented a statistical analysis of the least mean square (LMS) algorithm when a zero-memory saturation follows the adaptive filter output. The saturation nonlinearity was modeled by a scaled error function. This structure can model nonlinear effects in active noise and active vibration control systems when transducers are driven by large amplitude signals. This problem was first characterized as a nonlinear signal estimation problem. The resulting mean-square error (MSE) performance surface was studied in detail. New analytical expressions were obtained for the optimum weight vector and for the minimum achievable MSE as functions of the

system's degree of nonlinearity. The new results were shown to be useful for adaptive algorithm design and evaluation. The LMS algorithm analysis with a nonlinearity in the adaptation loop yielded deterministic nonlinear recursions for the mean weight and MSE behavior for Gaussian inputs and slow adaptation. A simplified model was obtained for the case of white inputs. Simple expressions for small step sizes have also been derived for the steady-state mean weight and MSE behavior. 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

PROOF THAT $\nabla^2\xi(\tilde{W})$ IS POSITIVE DEFINITE

The Hessian of $\xi(W)$ is given by (40), shown at the bottom of the page. At $\tilde{W} = cW^o$, (40) becomes (41), shown at the bottom of the page.

Assuming R_{xx} is positive definite, (41) can be written as

$$\nabla^2\xi(\tilde{W}) = R_{xx}^{1/2}\left\{aI + bR_{xx}^{1/2}W^o{W^o}^{\mathrm{T}}R_{xx}^{1/2}\right\}R_{xx}^{1/2} \tag{42}$$

where $R_{xx}^{1/2}$ is symmetric and nonsingular. Thus, (42) is of the form $C^{\mathrm{T}}MC$, where C is nonsingular. The following result is now used [22, p. 254].

If A is positive definite and C is nonsingular, then $C^{\mathrm{T}}AC$ is also positive definite.

Thus, if M is positive definite, so is the Hessian.

The eigenvectors of

$$M = aI + bR_{xx}^{1/2}W^o{W^o}^{\mathrm{T}}R_{xx}^{1/2} \tag{43}$$

are $R_{xx}^{1/2}W^o$ and $N-1$ vectors orthogonal to it. Thus, M has $N-1$ eigenvalues equal to a and one eigenvalue given by $\gamma = a + b{W^o}^{\mathrm{T}}R_{xx}W^o$. M will be positive definite if all these eigenvalues are positive.

From (41)

$$a = \frac{2c{W^o}^{\mathrm{T}}R_{xx}W^o}{\sigma^2\left(\frac{c^2}{\sigma^2}{W^o}^{\mathrm{T}}R_{xx}W^o+1\right)^{3/2}} + \frac{2}{\left(\frac{2c^2}{\sigma^2}{W^o}^{\mathrm{T}}R_{xx}W^o+1\right)^{1/2}\left(\frac{c^2}{\sigma^2}{W^o}^{\mathrm{T}}R_{xx}W^o+1\right)}. \tag{44}$$

Using (2) and (10) in (44) yields

$$a = \frac{2c\eta^2}{(c^2\eta^2+1)^{3/2}} + \frac{2}{(2c^2\eta^2+1)^{1/2}(c^2\eta^2+1)} = \frac{2}{c(c^2\eta^2+1)^{1/2}} > 0 \tag{45}$$

since c has already been shown to be positive. Using (41), (2) and (10) yields, after simple algebraic manipulations

$$\gamma = a + b{W^o}^{\mathrm{T}}R_{xx}W^o = \frac{4c^4\eta^4+4c^2\eta^2+2}{(c^2\eta^2+1)^2(2c^2\eta^2+1)^{3/2}} > 0 \tag{46}$$

which completes the proof that the Hessian is positive definite for any finite η^2.

$$\begin{aligned}\nabla^2\xi(W) = {} & \left[\frac{2}{\sigma^2\left(\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^{3/2}}\right]R_{xx}\left(WW^{o^{\mathrm{T}}}+W^oW^{\mathrm{T}}\right)R_{xx} \\ & + \left[\frac{2W^{o^{\mathrm{T}}}R_{xx}W}{\sigma^2\left(\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^{3/2}} + \frac{2}{\left(\frac{2}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^{1/2}\left(\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)}\right]R_{xx} \\ & - \left[\frac{6W^{o^{\mathrm{T}}}R_{xx}W}{\sigma^4\left(\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^{5/2}} + \frac{12W^{\mathrm{T}}R_{xx}W+8\sigma^2}{\sigma^4\left(\frac{2}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^{3/2}\left(\frac{1}{\sigma^2}W^{\mathrm{T}}R_{xx}W+1\right)^2}\right] \\ & \times R_{xx}WW^{\mathrm{T}}R_{xx}.\end{aligned} \tag{40}$$

$$\begin{aligned}\nabla^2\xi(\tilde{W}) = {} & \left[\frac{2cW^{o^{\mathrm{T}}}R_{xx}W^o}{\sigma^2\left(\frac{c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^{3/2}} + \frac{2}{\left(\frac{2c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^{1/2}\left(\frac{c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)}\right]R_{xx} \\ & + \left[\frac{4c}{\sigma^2\left(\frac{c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^{3/2}} - \frac{6c^3W^{o^{\mathrm{T}}}R_{xx}W^o}{\sigma^4\left(\frac{c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^{5/2}}\right. \\ & \left. - \frac{12c^4W^{o^{\mathrm{T}}}R_{xx}W^o+8c^2\sigma^2}{\sigma^4\left(\frac{2c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^{3/2}\left(\frac{c^2}{\sigma^2}W^{o^{\mathrm{T}}}R_{xx}W^o+1\right)^2}\right]R_{xx}W^oW^{o^{\mathrm{T}}}R_{xx} \\ = {} & aR_{xx} + bR_{xx}W^oW^{o^{\mathrm{T}}}R_{xx}.\end{aligned} \tag{41}$$

APPENDIX B

DERIVATION OF (28)

Post-multiplying (17) by its transpose and averaging on the data yields

$$
\begin{aligned}
E\left\{W(n+1)W^{\mathrm{T}}(n+1)|W(n)\right\} &= W(n)W^{\mathrm{T}}(n) \\
&\quad + \mu\left[W(n)W^{o^{\mathrm{T}}}\overbrace{E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(1)}\right] \\
&\quad + \mu\left[W(n)W^{o^{\mathrm{T}}}E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}\right]^{\mathrm{T}} \\
&\quad + \mu\left[W(n)E\left\{z(n)X^{\mathrm{T}}(n)|W(n)\right\}\right] \\
&\quad + \mu\left[W(n)\overbrace{E\left\{z(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(2)}\right]^{\mathrm{T}} \\
&\quad - \mu\left[E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\}W^{\mathrm{T}}(n)\right] \\
&\quad - \mu\left[\overbrace{E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\}}^{(3)}W^{\mathrm{T}}(n)\right]^{\mathrm{T}} \\
&\quad + \mu^2\left[\overbrace{E\left\{z^2(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(4)}\right] \\
&\quad + 2\mu^2\left[\overbrace{E\left\{z(n)W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(5)}\right] \\
&\quad + \mu^2\left[\overbrace{E\left\{X(n)X^{\mathrm{T}}(n)W^{o}W^{o^{\mathrm{T}}}X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(6)}\right] \\
&\quad - 2\mu^2\left[\overbrace{E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(7)}\right] \\
&\quad - 2\mu^2\left[\overbrace{E\left\{z(n)g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(8)}\right] \\
&\quad + \mu^2\left[\overbrace{E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]X(n)X^{\mathrm{T}}(n)|W(n)\right\}}^{(9)}\right]. \qquad (47)
\end{aligned}
$$

The expected values in (47) are now determined.

Expression 1: Assuming $W(n)$ and $X(n)$ statistically independent

$$E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\} \approx R_{xx} \qquad (48)$$

Expression 2:

$$E\left\{z(n)X^{\mathrm{T}}(n)|W(n)\right\} = 0 \qquad (49)$$

Expression 3: This was already evaluated in (20) for $W(n)$ and $X(n)$ statistically independent.

Expression 4: Assuming $W(n)$ and $X(n)$ statistically independent,

$$
\begin{aligned}
&E\left\{z^2(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = E\left\{z^2(n)|W(n)\right\}E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\} = \sigma_z^2 R_{xx}. \qquad (50)
\end{aligned}
$$

Expression 5:

$$
\begin{aligned}
&E\left\{z(n)W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = E\left\{z(n)|W(n)\right\}E\left\{W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = 0 \qquad (51)
\end{aligned}
$$

Expression 6: Assuming $W(n)$ and $X(n)$ statistically independent, this term can be evaluated using the moment factoring theorem [3] since $x(n)$ is Gaussian. After some simple mathematical manipulations, it can be easily shown that

$$
\begin{aligned}
&E\left\{X(n)X^{\mathrm{T}}(n)W^{o}W^{o^{\mathrm{T}}}X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = 2R_{xx}W^{o}W^{o^{\mathrm{T}}}R_{xx} + W^{o^{\mathrm{T}}}R_{xx}W^{o}R_{xx}. \qquad (52)
\end{aligned}
$$

Expression 7: This expectation contains a nonlinear term. It can be written as

$$
\begin{aligned}
&E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = E\left\{g(y_1)y_2Y_3Y_3^{\mathrm{T}}\right\} \qquad (53)
\end{aligned}
$$

where $y_1(n) = W^{\mathrm{T}}(n)X(n)$, $y_2(n) = W^{o^{\mathrm{T}}}X(n)$, and $Y_3(n) = X(n)$. Following the same approach used in [18] and expanding the results to the vector case, the higher order moments can be broken into combinations of second moments as follows:

$$
\begin{aligned}
E\left\{g(y_1)y_2Y_3Y_3^{\mathrm{T}}\right\} &= E\left\{y_1y_2\right\}E\left\{y_1Y_3\right\}E\left\{y_1Y_3^{\mathrm{T}}\right\}A(y_1) \\
&\quad + \left[E\left\{y_1y_2\right\}E\left\{Y_3Y_3^{\mathrm{T}}\right\} + E\left\{y_1Y_3\right\}\right. \\
&\quad \left.\times E\left\{y_2Y_3^{\mathrm{T}}\right\} + E\left\{y_2Y_3\right\}E\left\{y_1Y_3^{\mathrm{T}}\right\}\right] \\
&\quad \times \frac{E\left\{y_1g(y_1)\right\}}{E\left\{y_1^2\right\}} \qquad (54)
\end{aligned}
$$

where

$$A(y_1) = \frac{E\left\{y_1^3g(y_1)\right\}}{E\left\{y_1^2\right\}^3} - 3\frac{E\left\{y_1g(y_1)\right\}}{E\left\{y_1^2\right\}^2}. \qquad (55)$$

The second moments in (54) are given by (56), shown at the bottom of the next page.

Inserting (56) in (54) and (53) yields

$$
\begin{aligned}
&E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} \\
&\quad = W^{\mathrm{T}}(n)R_{xx}W^{o}R_{xx}W(n)W^{\mathrm{T}}(n) \\
&\qquad \times R_{xx}A\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) \\
&\qquad + \frac{W^{\mathrm{T}}(n)E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\}}{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2|W(n)\right\}} \\
&\qquad \cdot\left[W^{\mathrm{T}}(n)R_{xx}W^{o}R_{xx} + R_{xx}W(n)W^{o^{\mathrm{T}}}R_{xx}\right. \\
&\qquad \left. + R_{xx}W^{o}W^{\mathrm{T}}(n)R_{xx}\right] \qquad (57)
\end{aligned}
$$

where $A(\cdot)$ is defined in (55). Thus

$$A\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) = \frac{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^3 g\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\}}{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2|W(n)\right\}^3} - 3\frac{W^{\mathrm{T}}(n)E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)|W(n)\right\}}{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2|W(n)\right\}^2}. \quad (58)$$

The numerator of the last term in (58) follows directly from (20). To determine the numerator of the first term, direct integration leads to

$$E\{y^3 g(y)\} = \frac{E\{y^2\}^2}{\left(\frac{E\{y^2\}}{\sigma^2}+1\right)^{3/2}} + 2\frac{E\{y^2\}^2}{\left(\frac{E\{y^2\}}{\sigma^2}+1\right)^{1/2}}. \quad (59)$$

Making $y = W^{\mathrm{T}}(n)X(n)$ and $E\{y^2|W(n)\} = W^{\mathrm{T}}(n)R_{xx}W(n)$ in (59) and using (20), it follows that (58) simplifies to

$$A\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) = -\frac{1}{\sigma^2\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)^{3/2}}. \quad (60)$$

Substituting (20) and (60) in (57) and rearranging the terms yields (61), shown at the bottom of the page.

Expression 8:

$$E\left\{z(n)g\left[W^{\mathrm{T}}(n)X(n)\right]X(n)X^{\mathrm{T}}(n)|W(n)\right\} = 0. \quad (62)$$

Expression 9: Proceeding as in [18, App. A] and generalizing the results to the vector case, it can be shown that

$$E\left\{g^2(y_1)Y_2Y_2^{\mathrm{T}}\right\} = E\left\{Y_2Y_2^{\mathrm{T}}\right\}E\left\{g^2(y_1)\right\} + E\left\{y_1Y_2\right\}E\left\{y_1Y_2^{\mathrm{T}}\right\}B(y_1) \quad (63)$$

where

$$B(y_1) = \frac{1}{E\{y_1^2\}}\left[\frac{E\{y_1^2 g^2(y_1)\}}{E\{y_1^2\}} - E\{g^2(y_1)\}\right]. \quad (64)$$

Defining $y_1 = W^{\mathrm{T}}(n)X(n)$, $y_2 = X(n)$, and using (24) leads to

$$E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]X(n)X^{\mathrm{T}}(n)|W(n)\right\} = \sigma^2\arcsin\left(\frac{W^{\mathrm{T}}(n)R_{xx}W(n)}{W^{\mathrm{T}}(n)R_{xx}W(n)+\sigma^2}\right)R_{xx} + R_{xx}W(n)W^{\mathrm{T}}(n)R_{xx}B\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) \quad (65)$$

where

$$B\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) = \frac{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2 g^2\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\}}{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2|W(n)\right\}^2} - \frac{E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]|W(n)\right\}}{E\left\{\left(W^{\mathrm{T}}(n)X(n)\right)^2|W(n)\right\}}. \quad (66)$$

The numerator of the second term of (66) is given by (24). To evaluate the numerator of the first term, the results in [18, App. A] can also be used to show that

$$E\left\{g^2(y)y^2\right\} = E\left\{y^2\right\}\left\{\sigma^2\arcsin\left(\frac{E\{y^2\}}{E\{y^2\}+\sigma^2}\right) + \frac{2E\{y^2\}}{\left(\frac{E\{y^2\}}{\sigma^2}+1\right)\sqrt{\frac{2E\{y^2\}}{\sigma^2}+1}}\right\}. \quad (67)$$

$$\begin{cases} E\{y_1y_2\} = W^{\mathrm{T}}(n)E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}W^o = W^{\mathrm{T}}(n)R_{xx}W^o \\ E\left\{y_1Y_3\right\} = E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}W(n) = R_{xx}W(n) \\ E\left\{y_2Y_3\right\} = E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\}W^o = R_{xx}W^o \\ E\left\{Y_3Y_3^{\mathrm{T}}\right\} = E\left\{X(n)X^{\mathrm{T}}(n)|W(n)\right\} = R_{xx}. \end{cases} \quad (56)$$

$$E\left\{g\left[W^{\mathrm{T}}(n)X(n)\right]W^{o^{\mathrm{T}}}X(n)X(n)X^{\mathrm{T}}(n)|W(n)\right\} = \frac{W^{\mathrm{T}}(n)R_{xx}W^oR_{xx} + R_{xx}W(n)W^{o^{\mathrm{T}}}R_{xx} + R_{xx}W^oW^{\mathrm{T}}(n)R_{xx}}{\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)^{1/2}} - \frac{W^{\mathrm{T}}(n)R_{xx}W^oR_{xx}W(n)W^{\mathrm{T}}(n)R_{xx}}{\sigma^2\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)^{3/2}} \quad (61)$$

Using $y = W^{\mathrm{T}}(n)X(n)$ and $E\{y^2|W(n)\} = W^{\mathrm{T}}(n)R_{xx}W(n)$ in (67) and substituting the result in (66) yields

$$B\left(W^{\mathrm{T}}(n)X(n)|W(n)\right) = \frac{2}{\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)\sqrt{\frac{2}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}}. \tag{68}$$

Substituting (68) in (65) leads to

$$E\left\{g^2\left[W^{\mathrm{T}}(n)X(n)\right]X(n)X^{\mathrm{T}}(n)|W(n)\right\} = \sigma^2\arcsin\left(\frac{W^{\mathrm{T}}(n)R_{xx}W(n)}{W^{\mathrm{T}}(n)R_{xx}W(n)+\sigma^2}\right)R_{xx} + \frac{2}{\left(\frac{1}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1\right)\sqrt{\frac{2}{\sigma^2}W^{\mathrm{T}}(n)R_{xx}W(n)+1}} \times R_{xx}W(n)W^{\mathrm{T}}(n)R_{xx}. \tag{69}$$

Finally, substituting the results of Expressions 1 through 9 in (47) yields the recursion for the conditional weight correlation matrix given in (28).

REFERENCES

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

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1985.

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

[4] S. D. Snyder and N. Tanaka, "Active control of vibration using a neural network," *IEEE Trans. Neural Networks*, vol. 6, pp. 819–828, July 1995.

[5] 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.

[6] T. A. C. M. Claasen and W. F. G. Meckelenbrauker, "Comparisons of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 510–518, June 1981.

[7] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 578–586, Apr. 1982.

[8] N. J. Bershad, "On the optimum data nonlinearity in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 69–76, Feb. 1986.

[9] ——, "On error saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 440–452, Apr. 1988.

[10] ——, "On weight update saturation nonlinearities in LMS adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 623–630, Apr. 1990.

[11] E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1367–1374, Nov. 1990.

[12] S. C. Douglas and T. H. Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Processing*, vol. 42, pp. 1352–1365, June 1994.

[13] ——, "Stochastic gradient adaptation under general error criteria," *IEEE Trans. Signal Processing*, vol. 42, pp. 1335–1351, June 1994.

[14] S. Koike, "Convergence analysis of a data echo canceller with a stochastic gradient adaptive FIR filter using the sign algorithm," *IEEE Trans. Signal Processing*, vol. 43, pp. 2852–2862, Dec. 1995.

[15] J. C. M. Bermudez and N. J. Bershad, "A nonlinear analytical model for the quantized LMS algorithm—the arbitrary step size case," *IEEE Trans. Signal Processing*, vol. 44, pp. 1175–1183, May 1996.

[16] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[17] 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.

[18] 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.

[19] A. Feuer and R. Cristi, "On the optimal weight vector of a perceptron with gaussian data and arbitrary nonlinearity," *IEEE Trans. Signal Processing*, vol. 41, pp. 2257–2259, June 1993.

[20] 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.

[21] O. J. Tobias, J. C. M. Bermudez, N. J. Bershad, and R. Seara, "Mean weight behavior of the Filtered-X LMS algorithm," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, 1998, pp. 3545–3548.

[22] G. Strang, *Linear Algebra and its Applications*. New York: Academic, 1980.

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

He joined the Department of Electrical Engineering of Universidade Católica de Pelotas (UCPel), Pelotas, Brazil, in 1994. He is currently an Associate Professor of Electrical Engineering and a researcher with the Biomedical Engineering Group at UCPel. 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 the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1978, the M.Sc. degree from COPPE/UFRJ, in 1981, and the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 1985, all in electrical engineering.

He joined the Department of Electrical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 1985. 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.

Dr. Bermudez served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996. He is currently serving his second term as Associate Editor in the same area. He is presently a Member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

Neil J. Bershad (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 has been 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, 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 cancellation.

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.