Convergence Behavior of NLMS Algorithm for Gaussian Inputs: Solutions Using Generalized Abelian Integral Functions and Step Size Selection

S. C. Chan · Y. Zhou

Abstract This paper studies the mean and mean square convergence behaviors of the normalized least mean square (NLMS) algorithm with Gaussian inputs and additive white Gaussian noise. Using the Price's theorem and the framework proposed by Bershad in IEEE Transactions on Acoustics, Speech, and Signal Processing (1986, 1987), new expressions for the excess mean square error, stability bound and decoupled difference equations describing the mean and mean square convergence behaviors of the NLMS algorithm using the generalized Abelian integral functions are derived. These new expressions which closely resemble those of the LMS algorithm allow us to interpret the convergence performance of the NLMS algorithm in Gaussian environment. The theoretical analysis is in good agreement with the computer simulation results and it also gives new insight into step size selection.

Keywords Normalized least mean square · Convergence

S. C. Chan (✉) · Y. Zhou
Department of Electrical and Electronic Engineering,
The University of Hong Kong,
Pokfulam Road, Hongkong
e-mail: scchan@eee.hku.hk

Y. Zhou
e-mail: yizhou@eee.hku.hk

1 Introduction

Adaptive filters are frequently employed to handle filtering problems in which the statistics of the underlying signals are either unknown *a priori*, or in some cases slowly-varying. Many adaptive filtering algorithms have been proposed and they are usually variants of the well known least mean square (LMS) [1] and the recursive least squares (RLS) [2] algorithms. An important variant of LMS algorithm is the normalized least mean square (NLMS) algorithm [3, 4], where the step size is normalized with respect to the energy of the input vector. Due to the numerical stability and computational simplicity of the LMS and the NLMS algorithms, they have been widely used in various applications [5]. Their convergence performance analyses are also long standing research problems. The convergence behavior of the LMS algorithm for Gaussian input was thoroughly studied in the classical work of Widrow *et al*. [1], in which the concept of independence assumption was introduced. Other related studies of the LMS algorithm with independent Gaussian inputs include [6–8]. On the other hand, the NLMS algorithm generally possesses an improved convergence speed over the LMS algorithm, but its analysis is more complicated due to the step size normalization. In [9] and [10], the mean and mean square behaviors of the NLMS algorithm for Gaussian inputs were studied. Analysis for independent Gaussian inputs in [11] also revealed the advantage of the NLMS algorithm over the LMS algorithm. Due to the difficulties in evaluating the expectations involved in the difference equations for the mean weight-error vector and its covariance matrix, general closed-form expressions for

these equations and the excess mean square error (EMSE) are in general unavailable. Consequently, the works in [9, 10] only concentrated on certain special cases of eigenvalue distribution of the input autocorrelation matrix. In [12, 13], particular or simplified input data model was introduced to facilitate the performance analysis so that useful analytical expressions can still be derived. In [14], the averaging principle was invoked to simplify the expectations involved in the difference equations. Basically, the normalization term is assumed to vary slowly with respect to the input correlation term and the power of the input vector is assumed to have a chi-square distribution with L degrees of freedom. In [15], the difference equation was converted to a stochastic differential equation to simplify the analysis assuming a small step size. In [16], the normalization term mentioned above was recognized as an Abelian integral [17], which was explicitly integrated using a transformation approach into elementary functions.[1] Recently, Sayed *et al*. [18] proposed a unified framework for analyzing the convergence of adaptive filtering algorithms. It has been applied to different adaptive filtering algorithms with satisfactory results [19].

In this paper, the convergence behaviors of the NLMS algorithm with Gaussian input and additive noise are studied. Using the Price's theorem [20] and the framework in [9, 10], new decoupled difference equations describing the mean and mean square convergence behaviors of the NLMS algorithm using the generalized Abelian integral functions are derived. The final results closely resemble the classical results for LMS in [1]. Moreover, it is found that the normalization process will always increase the maximum convergence rate of the NLMS algorithms over their LMS counterparts if the eigenvalues of the input autocorrelation matrix are unequal. Using the new solution for the EMSE, the step size parameters are optimized for white inputs which agrees with the approach previously proposed in [21] using calculus of variations. The theoretical analysis and some new bounds for step size selection are validated using Monte Carlo simulations.

The rest of this paper is organized as follows: In Section 2, the NLMS algorithm is briefly reviewed. In Section 3, the proposed convergence performance analysis is presented. Simulation results are given in Section 4 and conclusions are drawn in Section 5.
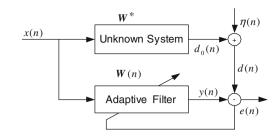
---

[1] To our best knowledge, the Abelian integrals which are also related to the exponential and elliptical integrals generally cannot be expressed in terms of a finite number of elementary functions. The approach in [16] seems to be an approximation.

Figure 1 Adaptive system identification.

2 NLMS Algorithm

Consider the adaptive system identification problem in Fig. 1 where an adaptive filter with coefficient or weight vector of order L, $\mathbf{W}(n) = [w_1(n), w_2(n), \ldots, w_L(n)]^T$, is used to model an unknown system with impulse response $\mathbf{W}^* = [w_1, w_2, \cdots, w_L]^T$. Here, $(\cdot)^T$ denotes the transpose of a vector or a matrix. The unknown system and the adaptive filter are simultaneously excited by the same input $x(n)$. The output of the unknown system $d_0(n)$ is assumed to be corrupted by a measurement noise $\eta(n)$ to form the desired signal $d(n)$ for the adaptive filter. The estimation error is given by $e(n) = d(n) - y(n)$. The NLMS algorithm under consideration assumes the following form:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \frac{\mu e(n) \mathbf{X}(n)}{\varepsilon + \alpha X^T(n) \mathbf{X}(n)}, \qquad (1)$$

where $\mathbf{X}(n) = [x(n), x(n-1), \cdots, x(n-L+1)]^T$ is the input vector at time instant n, μ is a positive step size constant to ensure convergence of the algorithm, and ε and α are positive constants. In the ε-NLMS algorithm [10], ε is a small positive value used to avoid division by zero and $\alpha = 1$. In the conventional LMS algorithm, $\varepsilon = 1$ and $\alpha = 0$. The above model can also be used to model the effect of prior knowledge of noise power by choosing ε to be $(1-\alpha)\widehat{\sigma}_x^2$, where $\widehat{\sigma}_x^2$ is some prior estimate of $E[\mathbf{X}^T(n)\mathbf{X}(n)]$ and α becomes a positive forgetting factor smaller than one.

3 Mean and Mean Square Convergence Analysis

To simplify the analysis, we assume that A1) the input signal $x(n)$ is a stationary ergodic process which is Gaussian distributed with zero mean and autocorrelation matrix $\mathbf{R}_{XX} = E[\mathbf{X}(n)\mathbf{X}^T(n)]$, A2) $\eta(n)$ is white Gaussian distributed with zero mean and variance σ_g^2, and A3) the well-known independent assumption [1] where $\mathbf{W}(n)$, $x(n)$ and $\eta(n)$ are considered statistically independent. Moreover, we denote $\mathbf{W}^* = \mathbf{R}_{XX}^{-1}\mathbf{P}_{dX}$ as the optimal Wiener solution, where $\mathbf{P}_{dX} = E[d(n)\mathbf{X}(n)]$ is the ensemble-averaged cross-correlation vector between $\mathbf{X}(n)$ and $d(n)$.

3.1 Mean Behavior

From (1), the update equation for the weight-error vector $\mathbf{v}(n) = \mathbf{W}^* - \mathbf{W}(n)$ is given by:

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \frac{\mu e(n)\mathbf{X}(n)}{\varepsilon + \alpha \mathbf{X}^T(n)\mathbf{X}(n)}. \qquad (2)$$

Taking expectation on both sides of (2), we get

$$E[\mathbf{v}(n+1)] = E[\mathbf{v}(n)] - \mu \mathbf{H}_1, \qquad (3)$$

where $\mathbf{H}_1 = E\left[e(n)\mathbf{X}(n)/\left(\varepsilon + \alpha \mathbf{X}^T(n)\mathbf{X}(n)\right)\right]$ and $E[\cdot]$ denotes the expectation over $\{\mathbf{v}(n), \mathbf{X}(n), \eta(n)\}$ and is written more clearly as $E_{\{v,X,\eta\}}[\cdot]$. Since $\mathbf{X}(n)$ and $\eta(n)$ are stationary, we can drop the time index n in the expectation to get $\mathbf{H}_1 = E\left[e\mathbf{X}/\left(\varepsilon + \alpha \mathbf{X}^T\mathbf{X}\right)\right]$. Using the independence assumption **A3**, we further have $\mathbf{H}_1 = E_{\{v\}}[\mathbf{H}]$, where $\mathbf{H} = E_{\{X,\eta\}}\left[e\mathbf{X}/(\varepsilon + \alpha \mathbf{X}^T\mathbf{X}) \,|\mathbf{v}\right]$.

In the conventional NLMS algorithm studied in [9] and [10] with $\alpha = 1$, similar difference equation for the mean weight-error vector (c.f. [10, Eq. 11])[2] was obtained:

$$E[\mathbf{v}(n+1)] = (\mathbf{I} - \mu \mathbf{F}_\varepsilon)E[\mathbf{v}(n)], \qquad (4)$$

where $\mathbf{F}_\varepsilon = E\left[\mathbf{X}(n)\mathbf{X}^T(n)/\left(\varepsilon + \mathbf{X}^T(n)\mathbf{X}(n)\right)\right]$ and $\mathbf{I}$ is the identity matrix. Moreover, $\mathbf{F}_\varepsilon$ was further diagonalized into $\mathbf{H}_\varepsilon$ whose i-th element is $[\mathbf{H}_\varepsilon]_{i,i} = \int_0^\infty \frac{\exp(-\beta\varepsilon)}{|\mathbf{I}+2\beta\mathbf{R}_{XX}|^{1/2}} \frac{\lambda_i}{1+2\beta\lambda_i} d\beta$ (c. f. [10, Eq. 14]), where λ_i is the i-th eigenvalue of $\mathbf{R}_{XX}$. It was evaluated analytically in [9] for three important cases with different eigenvalue distributions (in [10], only the first case was elaborated): (1) white input signal with $\lambda_1 = ... = \lambda_L$; (2) two signal subspaces with equal powers $\lambda_1 = ... = \lambda_K = a$ and $\lambda_{K+1} = ... = \lambda_L = b$; (3) distinct pairs, $\lambda_1 = \lambda_2$, $\lambda_3 = \lambda_4$, ..., $\lambda_{L-1} = \lambda_L$ (Assume L even). Besides these three special cases, no general solution to $\mathbf{H}_\varepsilon$ was provided. Therefore, general closed-form formulas for modeling the mean and mean square behaviors of the NLMS algorithm were unavailable in [9] and [10].

Here, we pursue another direction by treating some of these integrals as special functions and carry them throughout the analysis. The final formulas containing these special integral functions still allow us to clearly interpret the convergence behavior of the NLMS algorithm and determine appropriate step size parameters. More precisely, using Price's theorem [20] and the approach in [9, 10], it is shown in Appendix A that:

$$\mathbf{H}_1 = \mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T E[\mathbf{v}(n)] \qquad (5)$$

where $\mathbf{R}_{XX} = \mathbf{U}\Lambda\mathbf{U}^T$ is the eigenvalue decomposition of $\mathbf{R}_{XX}$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_L)$ contains the corresponding

eigenvalues. $\mathbf{D}_\Lambda$ is a diagonal matrix with the i-th diagonal entry given by (A-6):

$$[\mathbf{D}_\Lambda]_{i,i} = I_i(\Lambda) = \int_0^\infty \exp(-\beta\varepsilon)\left[\prod_{k=1}^L (2\alpha\beta\lambda_k+1)^{-1/2}\right](2\alpha\beta\lambda_i+1)^{-1}d\beta,$$

which is a generalized Abelian integral function where the conventional Abelian integral has the form $I_a(x) = \int_0^x [q(\beta)]^{-1/2}d\beta$ with $q(\beta)$ being a polynomial in β. It is also similar to $[\mathbf{H}_\varepsilon]_{i,i}$ in [10].

Substituting (5) into (3), the following difference equation for the mean weight-error vectors $E[\mathbf{v}(n+1)]$ and $E[\mathbf{v}(n)]$ is obtained

$$E[\mathbf{v}(n+1)] = \left(\mathbf{I} - \mu\mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T\right)E[\mathbf{v}(n)] \qquad (6)$$

(6) can also be written in the natural coordinate $\mathbf{V}(n) = \mathbf{U}^T\mathbf{v}(n)$ as

$$E[\mathbf{V}(n+1)] = (\mathbf{I} - \mu\Lambda\mathbf{D}_\Lambda)E[\mathbf{V}(n)], \qquad (7)$$

which is equivalent to L scalar first order finite difference equations as follows:

$$E[\mathbf{V}(n+1)]_i = (1 - \mu\lambda_i I_i(\Lambda))E[\mathbf{V}(n)]_i, \qquad (8)$$

where $E[V(n)]_i$ is the i-th element of the vector $E[V(n)]$ for $i = 1, 2, \cdots, L$.

For conventional LMS and NLMS algorithms, the above result agrees with the mean convergence result of the conventional NLMS algorithm in [10, Eq. 13], except that the effect of normalization is now more apparent. If all $I_i(\Lambda)$'s are equal to one, the analysis reduces to its unnormalized, conventional LMS counterpart. Therefore, the mean weight vector of the NLMS algorithm will converge if $|1 - \mu\lambda_i I_i(\Lambda)| < 1$, for all i, and the corresponding step size satisfies $\mu < 2/(\lambda_i I_i(\Lambda))$, for all i. To determine the maximum possible step size $\mu_{\max}$, we need to examine the maximum of the product $\lambda_i I_i(\Lambda)$:

$$\lambda_i I_i(\Lambda) = \lambda_i \int_0^\infty exp(-\beta\varepsilon)\left[\prod_{k=1}^L (2\alpha\beta\lambda_k+1)^{-1/2}\right](2\alpha\beta\lambda_i+1)^{-1}d\beta$$
$$= \int_0^\infty exp(-\beta\varepsilon)\left[\prod_{k=1}^L (2\alpha\beta\lambda_k+1)^{-1/2}\right]\left[2\alpha\beta + (\lambda_i)^{-1}\right]^{-1}d\beta. \qquad (9)$$

Since the factor $exp(-\beta\varepsilon)\left[\prod_{k=1}^L (2\alpha\beta\lambda_k+1)^{-1/2}\right]$ is common for all the products and it is positive, the maximum value in (9) also occurs at the largest eigenvalue $\lambda_{\max}$ with the corresponding value of $I_i(\Lambda)$ given by $I_{i\_\lambda_{\max}}(\Lambda)$. Therefore,

$$\mu < 2/(\lambda_{\max}I_{i\_\lambda_{\max}}(\Lambda)) = \mu_{\max} \qquad (10)$$

As a result, compared with the LMS algorithm, the maximum step size of the NLMS algorithm is scaled by a factor $1/I_{i\_\lambda_{\max}}(\Lambda)$. The fastest convergence rate of the algorithm occurs when $\mu = \mu_{\max}$ and it is limited by the mode corresponding to the smallest eigenvalue $\lambda_{\min}$ with

---

[2] For convenience, the variables are renamed according to the notation in this paper.

the corresponding value of $I_i(\Lambda)$ given by $I_{i\_\lambda_{\min}}(\Lambda)$, that is $1 - \lambda_{\min}I_{i\_\lambda_{\min}}(\Lambda)/(\lambda_{\max}I_{i\_\lambda_{\max}}(\Lambda))$. The smaller this value, the faster the convergence rate will be. From the definition of $I_i(\Lambda)$, it can be shown that $I_{i\_\lambda_{\min}}(\Lambda)/I_{i\_\lambda_{\max}}(\Lambda) \geq 1$. In other words, the eigenvalue spread $\lambda_{\max}/\lambda_{\min}$ is reduced by a factor $I_{i\_\lambda_{\max}}(\Lambda)/I_{i\_\lambda_{\min}}(\Lambda)$ after the normalization. Therefore, under the stated assumptions, the maximum convergence rate of the NLMS algorithm will always be faster than the LMS algorithm if the eigenvalues are unequal.

3.2 Mean Square Behavior

Post-multiplying (2) by its transpose and taking expectation gives

$$\Xi(n + 1) = \Xi(n) - \mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_3, \tag{11}$$

where $\Xi(n) = E[\mathbf{v}(n)\mathbf{v}^T(n)], \mathbf{M}_1 = \mu \mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T\Xi(n)$, $\mathbf{M}_2 = \mu\Xi(n)\mathbf{U}\mathbf{D}_\Lambda\Lambda\mathbf{U}^T$, $\mathbf{M}_3 = \mu^2E\left[e^2\mathbf{X}\mathbf{X}^T/(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})^2\right] = E_{\{\mathbf{v}\}}[\mathbf{s}_3]$, where $\mathbf{s}_3 = E_{\{\mathbf{X},\eta\}}\left[e^2\mathbf{X}\mathbf{X}^T/(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})^2|\mathbf{v}\right]$. Here, we have used the previous result in (5) to evaluate $\mathbf{M}_1$ and $\mathbf{M}_2$. $\mathbf{M}_3$ is evaluated in Appendix B to be

$$\mathbf{M}_3 = 2\mu^2\mathbf{U}\left\{\Lambda\left[\left(\mathbf{U}^T\Xi(n)\mathbf{U}\right)\circ\mathbf{I}(\Lambda)\right]\Lambda\right\}\mathbf{U}^T + \mu^2\mathbf{U}\overline{\mathbf{D}}_2\mathbf{U}^T \\ + \mu^2\sigma_g^2\mathbf{U}\Lambda\mathbf{I}'(\Lambda)\mathbf{U}^T, \tag{12}$$

where the diagonal matrix $\overline{\mathbf{D}}_2$ results from (B-7) and its i-th element $\left[\overline{\mathbf{D}}_2\right]_{i,i} = \sum_k \lambda_k\lambda_iI_{ki}(\Lambda)\left[\mathbf{U}^T\Xi(n)\mathbf{U}\right]_{k,k}$, $\circ$ denotes element-wise product of two matrices (Hadamard product), $\mathbf{I}(\Lambda)$ and $\mathbf{I}'(\Lambda)$ are defined in (B-5) and (B-8) and their elements are also generalized Abelian integral functions. Substituting (12) into (11) gives

$$\Xi(n + 1) = \Xi(n) - \mu\mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T\Xi(n) - \mu\Xi(n)\mathbf{U}\mathbf{D}_\Lambda\Lambda\mathbf{U}^T \\ + 2\mu^2\mathbf{U}\left\{\Lambda\left[\left(\mathbf{U}^T\Xi(n)\mathbf{U}\right)\circ\mathbf{I}(\Lambda)\right]\Lambda\right\}\mathbf{U}^T + \mu^2\mathbf{U}\overline{\mathbf{D}}_2\mathbf{U}^T \\ + \mu^2\sigma_g^2\mathbf{U}\Lambda\mathbf{I}'(\Lambda)\mathbf{U}^T. \tag{13}$$

(13) can be further simplified in the natural coordinate by pre- and post-multiplying $\Xi(n)$ by $\mathbf{U}^T$ and $\mathbf{U}$ to give:

$$\Phi(n + 1) = \Phi(n) - \mu\Lambda\mathbf{D}_\Lambda\Phi(n) - \mu\Phi(n)\mathbf{D}_\Lambda\Lambda \\ + 2\mu^2[\Lambda(\Phi(n)\circ\mathbf{I}(\Lambda))\Lambda] + \mu^2\widetilde{\mathbf{D}}_2 + \mu^2\sigma_g^2\Lambda\mathbf{I}'(\Lambda) \tag{14}$$

where $\Phi(n) = \mathbf{U}^T\Xi(n)\mathbf{U}$ and $\left[\widetilde{\mathbf{D}}_2\right]_{i,i} = \sum_k \lambda_k\lambda_iI_{ki}(\Lambda)[\Phi(n)]_{k,k}$. From (14), we can get the i-th diagonal value of $\Phi(n)$ as follows

$$\Phi_{i,i}(n + 1) = \Phi_{i,i}(n) - 2\mu I_i(\Lambda)\lambda_i\Phi_{i,i}(n) \\ + 2\mu^2\lambda_i^2I_{ii}(\Lambda)\Phi_{i,i}(n) + \mu^2\sum_k \lambda_k\lambda_iI_{ki}(\Lambda)\Phi_{k,k}(n) \\ + \mu^2\sigma_g^2\lambda_iI_i'(\Lambda). \tag{15}$$

From numerical results, the term $\mu^2\sum_k \lambda_k\lambda_iI_{ki}(\Lambda)\Phi_{k,k}(n)$ is very small for small EMSE and (15) can be approximated as

$$\Phi_{i,i}(n + 1) \approx \left(1 - 2\mu I_i(\Lambda)\lambda_i + 2\mu^2I_{ii}(\Lambda)\lambda_i^2\right)\Phi_{i,i}(n) + \mu^2\sigma_g^2\lambda_iI_i'(\Lambda). \tag{16}$$

To study the step size for mean squares convergence of the algorithm, we first assume that the algorithm converges and then determine an upper bound of the EMSE at the steady state. From this expression, we are able to find the step size bound for a finite EMSE and hence mean square convergence. As we shall see below, this step size bound depends weakly on the signals. Alternatively, we can find an approximate signal independent upper bound for small EMSE by ignoring the term $\mu^2\sum_k \lambda_k\lambda_iI_{ki}(\Lambda)\Phi_{k,k}(n)$ since it is very small for small EMSE. Consequently, (16) suggests that the algorithm will converge in the mean squares sense when $|1 - 2\mu I_i(\Lambda)\lambda_i + 2\mu^2I_{ii}(\Lambda)\lambda_i^2| < 1$. This gives $\mu < I_i(\Lambda)/(\lambda_iI_{ii}(\Lambda))$ for all i. From the definitions of $I_i(\Lambda)$ and $I_{ii}(\Lambda)$, we have $I_i(\Lambda)/(\lambda_iI_{ii}(\Lambda)) = 2/\left(1 - I_{ii}''(\Lambda)/I_i(\Lambda)\right) > 2$ for $\alpha = 1$, where $I_i''(\Lambda) = \int_0^\infty \exp(-\beta\varepsilon)\left[\prod_{k=1}^L(2\beta\lambda_k + 1)^{-1/2}\right](2\beta\lambda_i + 1)^{-2}d\beta$. Therefore, a conservative signal independent stability bound for small EMSE is $\mu < 2$.

For a more precise upper bound, we first note that the EMSE at time instant n is given by $\mathrm{EMSE}(n) = \mathrm{Tr}(\Xi(n)\mathbf{R}_{XX}) = \mathrm{Tr}(\Phi(n)\Lambda)$. Assuming that algorithm converges, it is shown in Appendix B that the last two terms on the right hand side of (15) is upper bounded by $\mu^2\sigma_e^2(\infty)\lambda_iI_i'(\Lambda)$ at the steady state. Hence, the steady state EMSE of the NLMS algorithm is approximately given by

$$\xi_{\mathrm{NLMS}}(\infty) = \mathrm{Tr}(\Phi(\infty)\Lambda) \approx \tfrac{1}{2}\mu\sigma_e^2(\infty)\phi_{\mathrm{NLMS}}, \tag{17}$$

where $\phi_{\mathrm{NLMS}} = \sum_{i=1}^L \frac{\lambda_iI_i'(\Lambda)}{I_i(\Lambda) - \mu\lambda_iI_{ii}(\Lambda)}$. Using the fact that $\sigma_e^2(\infty) = \xi_{\mathrm{NLMS}}(\infty) + \sigma_g^2$, one gets

$$\xi_{\mathrm{NLMS}}(\infty) \approx \frac{\tfrac{1}{2}\mu\sigma_g^2\phi_{\mathrm{NLMS}}}{1 - \tfrac{1}{2}\mu\phi_{\mathrm{NLMS}}}. \tag{18}$$

It can be seen that $\xi_{\mathrm{NLMS}}(\infty)$ is unbounded when either its denominator becomes zero or when ϕ_{NLMS} becomes unbound when any of the denominators of its partial sum becomes zero. This gives respectively the following two conditions:

$$\mu < 2/\phi_{\mathrm{NLMS}}, \tag{19a}$$

$$0 < \mu < I_i(\Lambda)/\lambda_iI_{ii}(\Lambda). \tag{19b}$$

For the LMS case, $I_{ii}(\Lambda) = I_i(\Lambda) = I_i'(\Lambda) = 1$ and the corresponding conditions are:

$$\mu < 2/\phi_{\text{LMS}}, \tag{20a}$$

$$0 < \mu < 1/\lambda_i, \tag{20b}$$

where $\phi_{\text{LMS}} = \sum_{i=1}^{L} \frac{\lambda_i}{1 - \mu \lambda_i}$. (20a) and (20b) are identical to the necessary and sufficient conditions for the mean square convergence of the LMS algorithm previously obtained in [6]. Similar results are obtained in [7] by solving the difference equation in $\Phi(n)$ and in [8] by a matrix analysis technique.

In [7], a lower bound of the maximum step size is also obtained. Using a similar approach, we now derive a step size bound for the NLMS algorithm. First we rewrite (19a) as:

$$\sum_{i=1}^{L} \frac{\mu \lambda_i c_i}{1 - \mu \lambda_i d_i} < 2, \tag{21}$$

where $c_i = I_i'(\Lambda)/I_i(\Lambda)$ and $d_i = I_{ii}(\Lambda)/I_i(\Lambda)$. Let $u = 2\,\mu^{-1}$ and rewrite (21) as

$$\ell(u) - \sum_{i=1}^{L} \lambda_i c_i l_i(u) = \prod_{i=1}^{L} (u - \bar{u}_i) = \sum_{i=0}^{L} (-1)^{L-i} b_{L-i} u^i = 0, \tag{22}$$

where, $\ell(u) = \prod_{i=1}^{L} (u - 2\lambda_i d_i)$, $l_i(u) = \ell(u)/(u - 2\lambda_i d_i)$, and $\bar{u}_i^{-1}$ are the roots of (22). The largest root of (22) (smallest root of (21)) is upper bounded (lower bounded) by [26]

$$u_{N\_\max} \leq \frac{1}{L} \left(s_1 + \sqrt{(L-1)(Ls_2 - s_1^2)} \right) \tag{23}$$

where $s_1 = \sum_{i=1}^{L} \bar{u}_i = b_1$ and $s_2 = \sum_{i=1}^{L} \bar{u}_i^2 = b_1^2 - 2b_2$. By comparing the coefficients on different sides of (22), one also gets

$$b_1 = \sum_{i=1}^{L} \lambda_i (2d_i + c_i)$$

and

$$b_2 = 4 \sum_{1 \leq i \neq j \leq L} \lambda_i \lambda_j d_i d_j + \sum_{i=1}^{L} \lambda_i c_i \left(2 \sum_{1 \leq j \neq i \leq L} \lambda_j d_j \right). \tag{24}$$

From (23), a more convenient lower bound of $\mu_{\max}$ can be obtained as follows

$$\mu_{\max} \geq \frac{2L}{b_1 + \sqrt{(L-1)^2 b_1^2 - 2L(L-1)b_2}} \geq \frac{2L}{b_1 + \sqrt{(L-1)^2 b_1^2}}$$

$$= \frac{2}{b_1} = \frac{2}{Tr\left(\Lambda(\mathbf{I}'(\Lambda) + 2\,diag(\mathbf{I}(\Lambda)))\mathbf{D}_\Lambda^{-1} \right)}. \tag{25}$$

where $diag(\mathbf{I}(\Lambda))$ is a diagonal matrix with the i-th diagonal value equal to $I_{ii}(\Lambda)$.

From simulation results, we found that ϕ_{NLMS} is rather close to one. Hence, $\mu < 2$ is a very useful rule of thumb estimate of the step size bound, since it does not depend on any prior knowledge of the Gaussian inputs.

Comparing (13) and the result in [10, Eq. 22], it can be found that they are identical except that all the integrals in [10] are coupled in their original forms. In contrast, the decoupled forms in terms of the generalized Abelian integral functions in (13)–(18) obtained with the proposed approach are very similar to those of the LMS algorithm. Moreover, we are also able to express the stability bound and EMSE in terms of these special functions, which are new to our best knowledge. When $I_i(\Lambda)$, $I_i'(\Lambda)$ and $I_{ii}(\Lambda)$ are equal to one, our analysis will reduce to the classical results of the LMS algorithm. Next, we shall make use of these analytical expressions for step size selection.

3.3 Step Size Selection

For white input, $I(\Lambda)$ and $I'(\Lambda)$ will reduce respectively to $\frac{\exp(\varepsilon/2\lambda)}{2\lambda} E_{L/2+1}\left(\frac{\varepsilon}{2\alpha\lambda}\right)$ and $\frac{\exp(\varepsilon/2\lambda)}{(2\alpha\lambda)^2}\left[E_{L/2}\left(\frac{\varepsilon}{2\alpha\lambda}\right) - E_{L/2+1}\left(\frac{\varepsilon}{2\alpha\lambda}\right) \right]$, where $E_n(x) = \int_1^\infty (\exp(-xt)/t^n) dt$ is the generalized exponential integral function ($E_{-n}(x)$ is also known as the Misra function). For small ε, one gets $E_n(\varepsilon/2\lambda) \approx 1/(n-1)$ for $n > 1$. In this case, the NLMS algorithm will have the same convergence rate of the LMS algorithm if $\mu_{\text{LMS}} = \mu \frac{\exp(\varepsilon/2\lambda)}{2\alpha\lambda} E_{L/2+1}\left(\frac{\varepsilon}{2\alpha\lambda}\right) \approx \mu \frac{1}{2\alpha\lambda(L/2)}$, or equivalently $\mu \approx \mu_{\text{LMS}} \alpha \lambda L$. For the maximum possible adaptation speed of the LMS algorithm, $\mu_{\text{LMS,opt}} = \frac{\lambda}{(L-1)^2 + E[x^4]} \approx 1/(\lambda L)$ for large L. As a result, $\mu \approx \alpha$ and one gets the following update

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \frac{e(n)\mathbf{X}(n)}{(\varepsilon/\alpha) + \mathbf{X}^T(n)\mathbf{X}(n)}, \tag{26}$$

which agrees with the optimum data nonlinearity for LMS adaptation in white Gaussian input obtained in [21] using calculus of variations. The MSE improvement of this NLMS algorithm over the conventional LMS algorithm was analyzed in detail in [21]. In general, one could set $\alpha = 1$ and vary μ between 0 and 1 with a small ε to achieve a given MSE or match a given convergence rate as above. From simulation results, we found that the EMSE of the NLMS algorithm varies slightly with the eigenvalues for a given $\text{Tr}(\mathbf{R}_{XX})$. For small μ, (18) suggests that the LMS algorithm is almost independent of the eigenvalue spread for a given $\text{Tr}(\mathbf{R}_{XX})$ ($\phi_{\text{LMS}} \approx \Sigma_{i=1}^{L} \lambda_i$). Therefore, the relationship between μ and μ_{LMS} for the white input case, i.e. $\mu \approx \mu_{\text{LMS}} \text{TR}(\mathbf{R}_{XX})$, can be used as a reasonable approximation for the colored case and $\alpha = 1$. The corresponding EMSE is approximately $\frac{1}{2}\mu_{\text{LMS}}\sigma_g^2 \text{Tr}(\mathbf{R}_{XX}) = \frac{1}{2}\mu\sigma_g^2$. From (17), $\phi_{\text{NLMS}} \approx \sum_{i=1}^{L} \lambda_i I_i'(\Lambda)/I_i(\Lambda)$, which can be shown to be independent of scaling of input for small ε. From simulation, we also found that the EMSE

of the NLMS will increase slightly with the eigenvalue spread. Hence, $\frac{1}{2}\mu\sigma_g^2$ represents a useful lower bound for estimating the EMSE of NLMS algorithm. It is very attractive because it does not require the knowledge of the eigenvalues or eigenvalue spread of $\mathbf{R}_{XX}$. The corresponding estimate of the misadjustment is then $\frac{1}{2}\mu$. A similar upper bound can be estimated empirically from simulation results to be presented.

4 Simulation Results

In this section we shall conduct computer experiments using both simulated and real world signals to verify the analytical results obtained in Section 3.

(1) Simulated signals

These simulations are carried out using the system identification model shown in Fig. 1. All the learning curves are obtained by averaging the results of $K=200$ independent runs. The unknown system to be estimated is a FIR filter with length L. Its weight vector $\mathbf{W}^*$ is randomly generated and normalized to unit energy. The input signal $x(n)=ax(n-1)+v(n)$ is a first order AR process, where $v(n)$ is a white Gaussian noise sequence with zero mean and variance σ_v^2. $0<\alpha<1$ is the correlation coefficient. The additive Gaussian noise $\eta(n)$ has zero mean and variance σ_g^2.

For the NLMS algorithm, we set $\alpha=1$, $\varepsilon=10^{-4}$ and vary μ. The values of the special integral functions, $I_i(\Lambda)$ in (A-6), $I_{ij}(\Lambda)$ in (B-5), and $I_i'(\Lambda)$ in (B-8) are evaluated numerically using the method introduced in [22]. For the mean convergence, the norm of the mean square weight-error vector is used as the performance measure:

$$||\mathbf{v}_A(n)||_2 = \sqrt{\sum_{i=1}^{L}\left[\frac{1}{K}\Sigma_{j=1}^{K}v_i^{(j)}(n)\right]^2}, i = 1,\cdots,L, \; j$$
$$= 1,\cdots,K,$$

where $v_i^{(j)}(n)$ is the i-th component of $\mathbf{v}(n)$ at time n in the j-th independent run. For the mean square convergence results, $\text{EMSE}(n) = \text{Tr}(\Phi(n)\Lambda)$, or the misadjustment $M(n) = \text{EMSE}(n)\big/\sigma_g^2$, is used as the performance measure. The theoretical results are computed from (8) and (16).

Two experiments are conducted. In the first experiment, we compare our analytical results with those in [14] (Eq. 14) and [16] (Eq. 25) for mean square convergence. Two filter lengths with $L=8$ and $L=24$ are evaluated for two cases: (1) $\mu=0.1$, $\sigma_g^2 = 10^{-5}$, and a less colored input with $\alpha=0.5$; (2) $\mu=0.1$, $\sigma_g^2 = 10^{-4}$, and a more colored input with $\alpha=0.9$. From Fig. 2 (a) and (b), it can be seen that when the input is less colored, all the approaches show
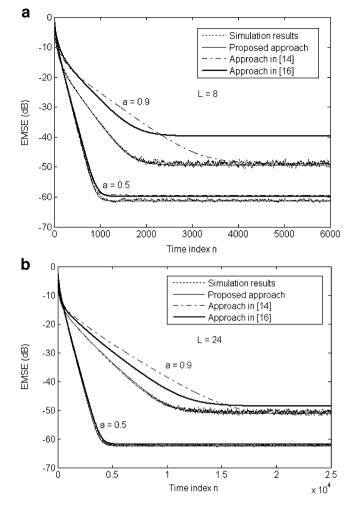

Figure 2 Comparison of the proposed analytical results with those in [14] and [16]. **a** L=8, **b** L=24.

good agreement with simulation results. When the input is very colored, our approach gives more accurate results than [14] and [16]. When L is small, there are considerable discrepancies between the theoretical and simulation results in [14] and [16]. The main reason is that both [14] and [16] assume that the denominator in (2) is uncorrelated with the numerator and an "average" but constant normalization for all the eigen-modes results. When the input is very colored, the scaling constants according to (8), $I_i(\Lambda)$, are considerably different for different modes. Hence, the averaging principle is less accurate in describing the convergence behavior.

In experiment 2, we conduct extensive experiments to further verify our analytical results. The parameters are summarized in Table 1. The results concerning mean and mean square convergence are plotted in Fig. 3 (a), (b) and Fig. 3 (c)–(f) respectively. From these figures, it can be seen that the theoretical analysis agrees closely with the simulation results for all cases tested. The estimated lower bound $\frac{1}{2}\mu$ for misadjustment M obtained in Section 3.3 is

Table 1 List of parameters in experiment 2 (χ: eigenvalue spread of $\mathbf{R}_{XX}$, CF: correction factor).

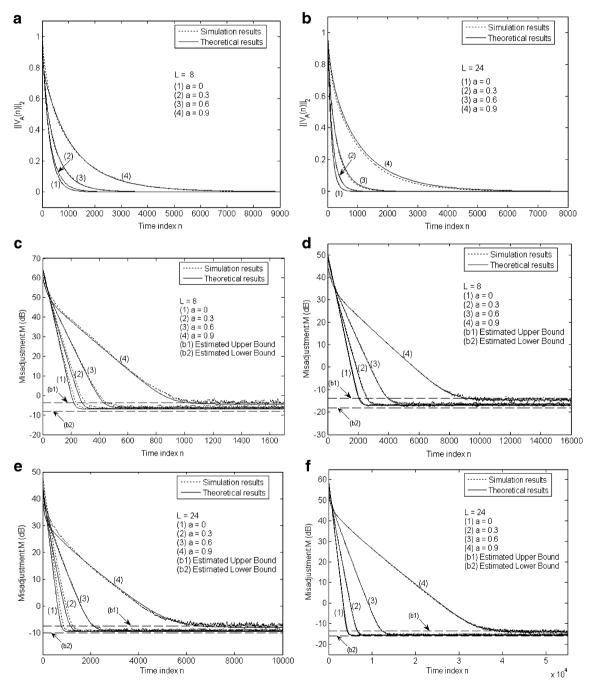| L	α(χ)	σ_g^2	σ_v^2	μ	CF	Fig.
8	0(1), 0.3(3.2), 0.6(12.2), 0.9(113.4)	10^{-7}	$1/[3(1-\alpha^2)]$	0.3	2.7	3(c)
		10^{-5}	$1/(1-\alpha)^2$	0.03	2.7	3(a), 3(d)
24	0(1), 0.3(3.4), 0.6(15.2), 0.9(232.7)	10^{-4}	$3/(1-\alpha)^2$	0.2	1.7	3(b), 3(e)
		10^{-6}	$1/(1-\alpha^2)$	0.05	1.7	3(f)

Figure 3 Verification of the proposed analytical results with parameter settings given in Table 1, **(a)**, **(b)** Mean convergence; **(c)**–**(f)** Mean square convergence.

also plotted. It is accurate for moderate filter lengths and serves as a useful bound for short filter lengths. As mentioned earlier, the steady state misadjustment M increases slightly with the eigenvalue spread of the input signal. Therefore, by introducing a correction factor CF, we can empirically estimate an upper bound of M as $CF \cdot \frac{1}{2}\mu$. This also serves as a reference for the selection of μ to achieve at least a given misadjustment for moderate eigenvalue spreads. This CF is found to decrease slightly as L increases. Other simulation results (omitted here due to page limitation) also give similar conclusion except when μ is near to one, where slightly increased discrepancies between theoretical and simulation results are observed due to the limitation of the independent assumption **A3**. In summary, the advantages of the NLMS algorithm over the LMS algorithm are its good performance in colored inputs and its ease in step size selection, which make it very attractive in speech processing and other applications with time-varying input signal level.

(2) Real speech signals

To illustrate further the properties of the LMS and NLMS algorithms, real speech signals are employed to evaluate their performances in an acoustic echo cancellation application. The speech signals used for testing are obtained with courtesy from the open source in [24]. Figure 4 (a) represents the signal of a sentence articulated by a female speaker "a little black plate on the floor" plus a white Gaussian noise. The sampling rate is 8 kHz. The echo path used has a length of 128 and is a real one given as $m_1(k)$ in the ITU-T recommendation G.168 [25]. The background noise $\eta(n)$ has a power of $\sigma_g^2 = 10^{-4}$. For simplicity, no double talk is assumed in this experiment.

Two sets of step sizes for the LMS and NLMS algorithms are employed: 1) $\mu_{LMS}=0.08$, $\mu_{NLMS} = 0.5$, and 2) $\mu_{LMS} = 0.02$ and $\mu_{NLMS} = 0.1$. These values are chosen so that when the two algorithms are excited by the additive noise (i.e. during nearly silent time period), both algorithms will give a similar MSE. The residual errors after echo cancellation are depicted in Fig. 4 (b) and (c), respectively. From Fig. 4 (b), we can see that due to the nonstationary nature of the speech signal and hence the unavailability a priori knowledge of the input statistics, the LMS algorithm with a fixed step size of 0.08 diverges and plots after time index 10000 are omitted. In contrast, the performance of the NLMS is rather satisfactory. At a smaller step size of 0.02, it can be seen from Fig. 4 (c) that LMS algorithm converges but its performance is severely degraded by the non-stationarity of the speech signal, whereas the NLMS algorithm again has a satisfactory performance. This is due to the rapidly changing input level and the colored nature of real speech signals.
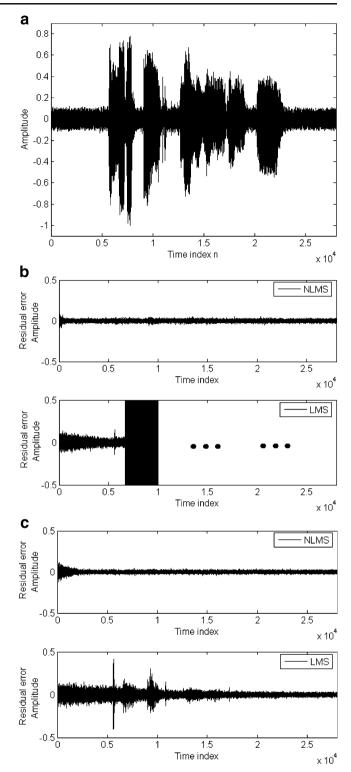

Figure 4 a A real speech signal of the sentence "a little black plate on the floor". b, c The residual error in the acoustic echo cancellation application using the LMS and NLMS algorithms with different step sizes.

5 Conclusions

A new convergence analysis of the NLMS algorithm using Price's theorem and the framework proposed in [9, 10] in Gaussian input and noise environment is presented. New expressions are derived for stability bound, steady state EMSE and decoupled difference equations describing the mean and mean square convergence behaviors of the NLMS algorithm using the generalized Abelian integral functions. The theoretical models are in good agreement with the simulation results and guidelines for step size selection are discussed.

Appendix

Appendix A: Evaluation of $\mathbf{H}_1$

As $\eta(n)$ and $x(n)$ are assumed to be statistically independent, and $\mathbf{X}$ are jointly Gaussian with autocorrelation matrix $\mathbf{R}_{XX}$, one gets

$$\mathbf{H} = C_R \iint_{L+1 fold} \frac{e\mathbf{X}}{\varepsilon + \alpha \mathbf{X}^T \mathbf{X}} \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{R}_{XX}^{-1} \mathbf{X}\right) f(\eta) d\eta d\mathbf{X}, \quad (A-1)$$

where $C_R = (2\pi)^{-L/2} |\mathbf{R}_{XX}|^{-1/2}$ and $f(\eta)$ is the probability density function (PDF) of the Gaussian noise η. $|\cdot|$ denotes the determinant of a matrix. Similar to [9] and [10], let us consider the integral

$$\mathbf{F}(\beta) = C_R \iint_{L+1 fold} \frac{e\mathbf{X} \exp\left(-\beta(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})\right)}{\varepsilon + \alpha \mathbf{X}^T \mathbf{X}} \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{R}_{XX}^{-1}\mathbf{X}\right) \quad (A-2)$$
$$\times f(\eta) d\eta d\mathbf{X}.$$

It can be seen that $\mathbf{H} = \mathbf{F}(0)$. Differentiating (A-2) w.r.t. β, one gets

$$\frac{d\mathbf{F}(\beta)}{d\beta} = -\exp(-\beta\varepsilon) C_R \iint_{L+1 fold} (e\mathbf{X}) \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{B}^{-1}(n)\mathbf{X}\right)$$
$$\times f(\eta) d\eta d\mathbf{X}, \quad (A-3)$$

where $\mathbf{B}(n) = \left(2\alpha\beta\mathbf{I} + \mathbf{R}_{XX}^{-1}\right)^{-1}$. For notation convenience, we shall simply write $\mathbf{B}$ for $\mathbf{B}(n)$. To evaluate the integral, we use the eigenvalue decomposition of $\mathbf{R}_{XX} = \mathbf{U}\Lambda\mathbf{U}^T$, where Λ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{R}_{XX}$ and U is a orthogonal matrix. The matrix B^{-1} can be written as $\mathbf{B}^{-1} = \mathbf{U}\left(2\alpha\beta\mathbf{I} + \Lambda^{-1}\right)\mathbf{U}^T = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T$, where $\mathbf{D}$ is a diagonal matrix with the i-th diagonal entry given by $d_{ii}(n) = \left(2\alpha\beta + \lambda_i^{-1}\right)^{-1}$. Noting that the determinants of $\mathbf{U}$ and $\mathbf{D}$ are respectively 1 and $|\mathbf{D}|^{-1/2} = \prod_{i=1}^{L}\left(2\alpha\beta + \lambda_i^{-1}\right)^{1/2}$, one can rewrite (A-3) as $d\mathbf{F}(\beta)/d\beta = -\gamma(\beta)\mathbf{L}_1$, where $C_B = (2\pi)^{-L/2} |\mathbf{B}|^{-1/2}$, $\mathbf{L}_1 = E_{\{\mathbf{X},\eta\}}[e\mathbf{X}|\mathbf{v}]|_{E[\mathbf{X}\mathbf{X}^T]=\mathbf{B}}$, and

$\gamma(\beta) = \exp(-\beta\varepsilon) \prod_{i=1}^{L}(2\alpha\beta\lambda_i + 1)^{-1/2}$. Using $e = \mathbf{X}^T\mathbf{v} + \eta$, we have $r_{x_i e} = E[x_i e|\mathbf{v}]|_{E[\mathbf{X}\mathbf{X}^T]=\mathbf{B}} = E\left[x_i(\mathbf{X}^T\mathbf{v} + \eta)|\mathbf{v}\right]|_{E[\mathbf{X}\mathbf{X}^T]=\mathbf{B}} = \mathbf{B}_i\mathbf{v}$, where $\mathbf{B}_i$ is the i-th row of $\mathbf{B}$. Finally, we have $\mathbf{L}_1 = \mathbf{B}\mathbf{v}(n)$.

Substituting it into (A-3) and integrating w.r.t. β yields

$$\mathbf{F}(\beta) = \mathbf{I}(\beta)\mathbf{v}(n), \quad (A-4)$$

where $\mathbf{I}(\beta) = -\int^{\beta}\gamma(\beta)\mathbf{B}d\beta$ and the constant of integration is equal to zero because of the boundary condition $\mathbf{F}(\infty) = 0$. To evaluate $\mathbf{I}(0)$ and hence $\mathbf{H} = \mathbf{F}(0) = \mathbf{I}(0)\mathbf{v}(n)$, we note that $\mathbf{B} = \mathbf{U}\mathbf{D}(n)\mathbf{U}^T$ and thus

$$\mathbf{I}(0) = \mathbf{U}\left(-\int^0 \gamma(\beta)\mathbf{D}(n)d\beta\right)\mathbf{U}^T = \mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T, \quad (A-5)$$

where $\mathbf{D}_\Lambda = \text{diag}(I_1(\Lambda), ..., I_L(\Lambda))$ and

$$I_i(\Lambda) = \int_0^\infty exp(-\beta\varepsilon)\left[\prod_{k=1}^{L}(2\alpha\beta\lambda_k + 1)^{-1/2}\right](2\alpha\beta\lambda_i + 1)^{-1}d\beta, \quad (A-6)$$
$$i = 1, 2, \cdots, L,$$

which can be numerically evaluated.

Finally, we have the desired result in (5)

$$\mathbf{H}_1 = E_{\{\mathbf{v}\}}[\mathbf{H}] = E_{\{\mathbf{v}\}}[\mathbf{F}(0)] = \mathbf{U}\Lambda\mathbf{D}_\Lambda\mathbf{U}^T E[\mathbf{v}(n)] \quad (A-7)$$

For the LMS algorithm, $\mathbf{B}$ will be equal to $\mathbf{R}_{XX}$ and accordingly $\mathbf{D}_\Lambda$ is equal to the identity matrix and $\sigma_e^2(\mathbf{v}) = \mathbf{v}^T\mathbf{R}_{XX}\mathbf{v} + \sigma_g^2$.

Appendix B: Evaluation of $\mathbf{s}_3$ and $\mathbf{M}_3$

As in Appendix A, we write $\mathbf{s}_3 = E_{\{\mathbf{X},\eta\}}\left[e^2\mathbf{X}\mathbf{X}^T \big/ (\varepsilon + \alpha\mathbf{X}^T\mathbf{X})^2 |\mathbf{v}\right]$ as:

$$\mathbf{s}_3 = C_R \iint_{L+1 fold} \frac{e^2\mathbf{X}\mathbf{X}^T}{(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})^2} exp\left(-\frac{1}{2}\mathbf{X}^T\mathbf{R}_{XX}^{-1}\mathbf{X}\right) f(\eta) d\eta d\mathbf{X}. \quad (B-1)$$

Let us define

$$\overline{\mathbf{F}}(\beta) = C_R \iint_{L+1 fold} \frac{e^2\mathbf{X}\mathbf{X}^T exp\left(-\beta(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})\right)}{(\varepsilon + \alpha\mathbf{X}^T\mathbf{X})^2} \quad (B-2)$$
$$\cdot exp\left(-\frac{1}{2}\mathbf{X}^T\mathbf{R}_{XX}^{-1}\mathbf{X}\right) f(\eta) d\eta d\mathbf{X}.$$

Comparing (B-2) with (B-1), it can be seen that $\mathbf{s}_3 = \overline{\mathbf{F}}(0)$. To evaluate $\overline{\mathbf{F}}(\beta)$, we differentiate (B-2) twice w.r.t. β to get $d^2\overline{\mathbf{F}}(\beta)/d\beta^2 = \gamma(\beta)\mathbf{L}_3$, where $\mathbf{L}_3 = E_{\{\mathbf{X},\eta\}}\left[e^2\mathbf{X}\mathbf{X}^T|\mathbf{v}\right]|_{E[\mathbf{X}\mathbf{X}^T]=\mathbf{B}}$, and $\gamma(\beta)$, C_B, and $\mathbf{B}$ have been defined in Appendix A. Consider the (i, j)-th element of $\mathbf{L}_3$: $\mathbf{L}_{3,i,j} = E_{\{\mathbf{X},\eta\}}\left[e^2 x_i x_j|\mathbf{v}\right]|_{\mathbf{B}}$. Using Price's theorem, we have $\partial \mathbf{L}_{3,i,j}/\partial r_{x_i x_j} = E_{\{\mathbf{X},\eta\}}[e^2|\mathbf{v}]|_{\mathbf{B}} = \sigma_e^2$, where $\sigma_e^2 = \mathbf{v}^T\mathbf{B}\mathbf{v} + \sigma_g^2$. Integrating $\partial\mathbf{L}_{3,i,j}/\partial r_{x_i x_j}$ w.r.t. $r_{x_i x_j}$ gives $\mathbf{L}_{3,i,j} = \sigma_e^2 r_{x_i x_j} + c_{i,j}$, where $c_{i,j} = E\left[e^2 x_i x_j\right]|_{r_{x_i x_j}=0}$ is the integration

constant. Using Price's theorem again, we have $\partial c_{i,j}/\partial r_{x_i e} = E\left[\frac{de^2}{de}x_j\right]\Big|_{r_{x_j e}=0} = E\left[\frac{d^2e^2}{de^2}\right]r_{x_j e} = 2r_{x_j e}$. Integrating once again, we get $c_{i,j} = 2r_{x_j e}r_{x_i e}$. Combining, we have $\mathbf{L}_{3,i,j} = \sigma_e^2 r_{x_i x_j} + 2r_{x_j e}r_{x_i e}$, and hence $\mathbf{L}_3 = \sigma_e^2\mathbf{B} + 2\mathbf{B}\mathbf{v}\mathbf{v}^T\mathbf{B}$. Substituting it into $d^2\overline{\mathbf{F}}(\beta)/d\beta^2 = \gamma(\beta)\mathbf{L}_3$ gives

$$d^2\overline{\mathbf{F}}(\beta)/d\beta^2 = \gamma(\beta)\sigma_e^2\mathbf{B} + 2\gamma(\beta)\mathbf{B}\mathbf{v}\mathbf{v}^T\mathbf{B}. \qquad (B-3)$$

Integrating (B-3) w.r.t. β and using $\sigma_e^2(\mathbf{v}) = \mathbf{v}^T\mathbf{B}\mathbf{v} + \sigma_g^2$ yields

$$\mathbf{s}_3 = \mathbf{I}_1 + \mathbf{I}_2 + \sigma_g^2\mathbf{I}_3,$$

where $\mathbf{I}_1 = 2\int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\mathbf{B}\mathbf{v}\mathbf{v}^T\mathbf{B}d\beta_2 d\beta_1$, $\mathbf{I}_2 = \int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)(\mathbf{v}^T\mathbf{B}\mathbf{v})\mathbf{B}d\beta_2 d\beta_1$, and $\mathbf{I}_3 = \int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\mathbf{B}d\beta_2 d\beta_1$ with the boundary conditions: $\overline{\mathbf{F}}(\infty) = 0$ and $\partial\overline{\mathbf{F}}(\beta)\big/_{\beta=\infty} = 0$. The evaluation of the integrals $\mathbf{I}_1$, $\mathbf{I}_2$, and $\mathbf{I}_3$ is detailed below. Note, $E_{\{\mathbf{v}\}}\left[\sigma_e^2\right] = E_{\{\mathbf{v}\}}[\mathbf{v}^T\mathbf{B}\mathbf{v}] + \sigma_g^2$ is upper bounded by $E_{\{\mathbf{v}\}}[\mathbf{v}^T\mathbf{R}_{XX}\mathbf{v}] + \sigma_g^2 = \sigma_e^2(n)$, which is the MSE at time instant n. To simplify the derivation of the worse case EMSE, the term $\mathbf{I}_2 + \sigma_g^2\mathbf{I}_3$ can be approximated by $\sigma_e^2(\infty)\mathbf{I}_3$.

Evaluation of $\mathbf{I}_1$:

Using the eigenvalue decomposition of $\mathbf{B}$, one gets

$$\mathbf{I}_1 = 2\int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\mathbf{U}\mathbf{D}(n)\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\mathbf{D}(n)\mathbf{U}^T d\beta_2 d\beta_1 = \mathbf{U}\mathbf{D}_1\mathbf{U}^T,$$
$$(B-4)$$

where $\mathbf{D}_1 = 2\int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\mathbf{D}(n)\mathbf{V}\mathbf{V}^T\mathbf{D}(n)d\beta_2 d\beta_1$ and $\mathbf{V} = \mathbf{U}^T\mathbf{v}$. The (i,j)-th element of $\mathbf{D}_1$ is $[\mathbf{D}_1]_{i,j} = 2[\mathbf{V}\mathbf{V}^T]_{i,j}I_{i,j}$, where $I_{i,j} = \int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)d_{ii}(n)d_{jj}(n)d\beta_2 d\beta_1 = \lambda_i\lambda_j I_{ij}(\Lambda)$, and

$$I_{ij}(\Lambda) = \int_0^\infty \beta\exp(-\beta\varepsilon)\left[\prod_{k=1}^L (2\alpha\beta\lambda_k+1)^{-1/2}\right](2\alpha\beta\lambda_i+1)^{-1} \quad (B-5)$$
$$\times(2\alpha\beta\lambda_j+1)^{-1}d\beta$$

Hence $\mathbf{I}_1 = \mathbf{U}\mathbf{D}_1\mathbf{U}^T$, where $\mathbf{D}_1 = 2\Lambda\left[\left(\mathbf{V}\mathbf{V}^T\right)\circ\mathbf{I}(\Lambda)\right]\Lambda$, $[\mathbf{I}(\Lambda)]_{i,j} = I_{ij}(\Lambda)$, and $\circ$ denotes element-wise product of two matrices. $I_{ij}(\Lambda)$ can be evaluated numerically.

Evaluation of $\mathbf{I}_2$:

$$\mathbf{I}_2 = \int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\text{Tr}\left(\mathbf{v}\mathbf{v}^T\mathbf{B}\right)\mathbf{B}d\beta_2 d\beta_1$$
$$= \mathbf{U}\left(\int_0^\infty \beta_2\gamma(\beta_2)\text{Tr}\left(\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\mathbf{D}\right)\mathbf{D}d\beta_2\right)\mathbf{U}^T \quad (B-6)$$
$$= \mathbf{U}\left(\int_0^\infty \beta_2\gamma(\beta_2)\sum_k d_{kk}\left[\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\right]_{k,k}\mathbf{D}d\beta_2\right)\mathbf{U}^T$$
$$= \mathbf{U}\mathbf{D}_2\mathbf{U}^T,$$

where we have used $\mathbf{B}^{-1} = \mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T$ and the property of trace operation $\text{Tr}(\cdot)$. $\mathbf{D}_2$ is a diagonal matrix and its i-th diagonal element is given by:

$$[\mathbf{D}_2]_{i,i} = \int_0^\infty \beta_2\gamma(\beta_2)\sum_k d_{kk}d_{ii}\left[\mathbf{U}^T\mathbf{v}\mathbf{v}^T\mathbf{U}\right]_{k,k}d\beta_2$$
$$= \sum_k\left(\int_0^\infty \beta_2\gamma(\beta_2)d_{kk}d_{ii}d\beta_2\right)\left[\mathbf{V}\mathbf{V}^T\right]_{k,k} \quad (B-7)$$
$$= \sum_k \lambda_k\lambda_i I_{ki}(\Lambda)\left[\mathbf{V}\mathbf{V}^T\right]_{k,k}$$

where $I_{ki}(\Lambda)$ has already been defined in (B-5).

Evaluation of $\mathbf{I}_3$:

Since $\mathbf{I}_3 = \int_0^\infty\int_{\beta_1}^\infty \gamma(\beta_2)\mathbf{U}\mathbf{D}(n)\mathbf{U}^T d\beta_2 d\beta_1 = \mathbf{U}\mathbf{D}_3(n)\mathbf{U}^T$, it suffices to evaluate the integral of the diagonal elements of $\mathbf{D}_3(n)$ as follows:

$$[\mathbf{D}_3(n)]_{i,i} = \int_0^\infty\int_{\beta_2}^\infty \gamma(\beta_2)d_{ii}(n)d\beta_2 d\beta_1$$
$$= \int_0^\infty\int_0^\infty \gamma(\beta_2)d_{ii}(n)d\beta_1 d\beta_2$$
$$= \lambda_i\int_0^\infty \beta\exp(-\beta\varepsilon)\left[\prod_{k=1}^L(2\alpha\beta\lambda_k+1)^{-1/2}\right](2\alpha\beta\lambda_i+1)^{-1}d\beta.$$

Hence, $\mathbf{D}_3(n) = \Lambda\mathbf{I}'(\Lambda)$, where $\mathbf{I}'(\Lambda)$ is a diagonal matrix with its i-th diagonal element (also can be numerically evaluated) given by

$$I_i'(\Lambda) = \int_0^\infty \beta\exp(-\beta\varepsilon)\left[\prod_{k=1}^L(2\alpha\beta\lambda_k+1)^{-1/2}\right](2\alpha\beta\lambda_i+1)^{-1}d\beta.$$
$$(B-8)$$

Finally, we have

$$\mathbf{s}_3 = \mathbf{U}\mathbf{D}_1\mathbf{U}^T + \mathbf{U}\mathbf{D}_2\mathbf{U}^T + \sigma_g^2\mathbf{U}\mathbf{D}_3\mathbf{U}^T$$
$$= 2\mathbf{U}\left\{\Lambda\left[\left(\mathbf{V}\mathbf{V}^T\right)\circ\mathbf{I}(\Lambda)\right]\Lambda\right\}\mathbf{U}^T + \mathbf{U}\mathbf{D}_2\mathbf{U}^T + \sigma_g^2\mathbf{U}\Lambda\mathbf{I}'(\Lambda)\mathbf{U}^T$$

and

$$\mathbf{M}_3 = E_{\{\mathbf{v}\}}[\mathbf{s}_3] = 2\mu^2\mathbf{U}\left\{\Lambda\left[\left(\mathbf{U}^T\Xi(n)\mathbf{U}\right)\circ\mathbf{I}(\Lambda)\right]\Lambda\right\}\mathbf{U}^T$$
$$+ \mu^2\mathbf{U}\overline{\mathbf{D}}_2\mathbf{U}^T + \mu^2\sigma_g^2\mathbf{U}\Lambda\mathbf{I}'(\Lambda)\mathbf{U}^T.$$
$$(B-9)$$

From numerical results, $\mathbf{I}_2$ is close to zero if the EMSE is small, and $\mathbf{s}_3 \approx \mathbf{I}_1 + \sigma_g^2\mathbf{I}_3$. For the LMS algorithm, the normalization term in (B-2) is missing and $\mathbf{L}_3$ can be obtained by replacing $\mathbf{B}$ above by $\mathbf{R}_{XX}$.
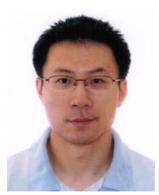
References

1. Widrow, B., McCool, J., Larimore, M. G., & Johnson, C. R., Jr. (1976). Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE, 64*, 1151–1162.
2. Plackett, R. L. (1972). The discovery of the method of least-squares. *Biometrika, 59*(2), 239–251.
3. Nagumo, J. I., & Noda, A. (1967). A learning method for system identification. *IEEE Transactions on Automatic Control, AC-12*, 282–287.

4. Bitmead, R. R., & Anderson, B. D. O. (1980). Performance of adaptive estimation algorithms in dependent random environments. *IEEE Transactions on Automatic Control, 25*, 788–794.

5. Haykin, S. (2001). *Adaptive filter theory* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

6. Horowitz, L. L., & Senne, K. D. (1981). Performance advantage of complex LMS for controlling narrowband adaptive arrays. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 29*(3), 722–736.

7. Feuer, A., & Weinstein, E. (1985). Convergence analysis of LMS filters with uncorrelated Gaussian data. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 33*(1), 222–230.

8. Foley, J. B., & Boland, F. M. (1988). A note on the convergence analysis of LMS adaptive filters with Gaussian data. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 36*(7), 1087–1089.

9. Bershad, N. J. (1986). Analysis of the normalized LMS algorithm with Gaussian inputs. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-34*(4), 793–806.

10. Bershad, N. J. (1987). Behavior of the ε-normalized LMS algorithm with Gaussian inputs. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35*(5), 636–644.

11. Tarrab, M., & Feuer, A. (1988). Convergence and performance analysis of the normalized LMS algorithm with uncorrelated Gaussian data. *IEEE Transactions on Information Theory, IT-34*(4), 680–691.

12. Rupp, M. (1993). The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes. *IEEE Transactions on Signal Processing, 41*(3), 1149–1160.

13. Slock, D. T. M. (1993). On the convergence behavior of the LMS and the normalized LMS algorithms. *IEEE Transactions on Signal Processing, 41*(9), 2811–2825.

14. Costa, M. H., & Bermudez, J. C. M. (2002). An improved model for the normalized LMS algorithm with Gaussian inputs and large number of coefficients. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2*, 1385–1388.

15. Barrault, G., Costa, M. H., Bermudez, J. C. M., & Lenzi, A. (2005). A new analytical model for the NLMS algorithm. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 4*, 41–44.

16. Lobato, E. M., Tobias, O. J., & Seara, R. (2006). Stochastic model for the NLMS algorithm with correlated Gaussian data. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 3*, 760–763.

17. Siegel, C. L. (1988). *Topics in complex function theory, vol. 2: Automorphic functions and abelian integrals*. NY: Wiley.

18. Yousef, N. R., & Sayed, A. H. (2001). A unified approach to the steady-state and tracking analyses of adaptive filters. *IEEE Transactions on Signal Processing, 49*(2), 314–324.

19. Sayed, A. H. (2003). *Fundamentals of adaptive filtering*. NY: Wiley.

20. Price, R. (1958). A useful theorem for nonlinear devices having Gaussian inputs. *IRE Transactions on Information Theory, IT-4*, 69–72.

21. Douglas, S. C., & Meng, T. H. Y. (1994). Normalized data nonlinearities for LMS adaptation. *IEEE Transactions on Signal Processing, 42*(6), 1352–1365.

22. Recktenwald, G. (2000). *Numerical methods with MATLAB: Implementations and applications*. Englewood Cliffs, NJ: Prentice-Hall.

23. Chan, S. C., & Zhou, Y. (2007). On the convergence analysis of the normalized LMS and the normalized least mean M-estimate algorithms. *In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*, 1059–1065.

24. Quatieri, T. F. (2002). *Discrete-time speech signal processing: Principles and practice*. Upper Saddle River, NJ: Prentice-Hall.

25. Digital Network Echo Cancellers. (2000). *ITU-T Recommendation G*. 168.

26. Jageman, O. L. (1975). Nonstationary blocking in telephone traffic. *Bell System Technical Journal, 54*(3). March.

S. C. Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from the University of Hong Kong in 1986 and 1992, respectively. He joined City Polytechnic of Hong Kong in 1990 as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, and is now a Professor. He held visiting positions in Microsoft Corporation, Redmond, USA, Microsoft Research Asia, University of Texas at Arlington, and Nanyang Technological University. Dr. Chan's main research area is signal processing and applications. His previous research covers fast transform algorithms, filter design and realization, multirate signal processing, adaptive and communications signal processing, data compression and image-based rendering. Dr. Chan has served in a number of professional and administrative committees. He is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, Associate Editors for IEEE Trans. Circuits and Systems I and Journal of Signal Processing systems, and was the Chairman of the IEEE HK Chapter of Signal Processing. He was the special session chairman of ICASSP2003 and is an organizing committee member of the IEEE 2010 ICIP.

Yi Zhou (S'05–M'07) was born in Chengdu, China. He received the B. Sc., M. Sc. degrees in 1997 and 1999 respectively, both in control engineering from Harbin Institute of Technology, Harbin, China, and Ph. D. degree in electrical engineering in 2006, from the University of Hong Kong, Hong Kong, China. He is currently a postdoctoral research fellow in the department of Electrical and Electronic Engineering, the University of Hong Kong. His research interests are adaptive signal processing, system identification, statistical signal processing and estimation theory. Dr. Zhou was the recipient of the Best Student Paper Award from the 13th IEEE Workshop on Statistical Signal Processing, 2005, Bordeaux, France.