Convergence Analysis of a Complex LMS Algorithm With Tonal Reference Signals

Mrityunjoy Chakraborty, *Senior Member, IEEE,* and Hideaki Sakai, *Senior Member, IEEE*

***Abstract*—Often one encounters the presence of tonal noise in many active noise control applications. Such noise, usually generated by periodic noise sources like rotating machines, is cancelled by synthesizing the so-called antinoise by a set of adaptive filters which are trained to model the noise generation mechanism. Performance of such noise cancellation schemes depends on, among other things, the convergence characteristics of the adaptive algorithm deployed. In this paper, we consider a multireference complex least mean square (LMS) algorithm that can be used to train a set of adaptive filters to counter an arbitrary number of periodic noise sources. A deterministic convergence analysis of the multireference algorithm is carried out and necessary as well as sufficient conditions for convergence are derived by exploiting the properties of the input correlation matrix and a related product matrix. It is also shown that under convergence condition, the energy of each error sequence is independent of the tonal frequencies. An optimal step size for fastest convergence is then worked out by minimizing the error energy.**

***Index Terms*—Active noise control (ANC), convergence analysis, least mean square (LMS) algorithm.**

I. Introduction

ACTIVE noise control (ANC) ([1], [2]) is an established procedure for cancellation of acoustic noise where first a model of the noise generation process is constructed, and then using that model, the so-called antinoise is synthesized that destructively interferes with the primary noise field and thus, minimizes its effect. Of the several applications of ANC studied so far, a case of particular interest is the cancellation of periodic noise containing multitonal components generated by rotating machines. Typical examples include propeller aircrafts, motorboats, vessels etc. that employ multiple engines, which may not remain synchronized perfectly and thus, may have varying speeds of rotation. The references ([3]–[5]) provide an account of major recent developments in this area. The active noise controller for such noise sources is illustrated in Fig. 1. There are p independent noise sources representing p rotating machines. Usually, the acoustic noise produced by each machine is narrowband, being dominated by a single tone frequency and is modeled as being generated by passing a sinusoid through a LTI filter. Synchronization signal, usually provided by a tachometer, provides information about the rotational speed of each machine and is used to generate sinusoidal reference signals. Each reference signal is filtered by an adaptive filter separately and the combined output of all the adaptive filters is used to drive a loudspeaker. The error microphone shown is used as an error sensing device to detect the noise difference between the primary noise and the secondary noise produced by the loudspeaker and to convert this differential noise into an electrical error signal which is used to adjust the adaptive filter weights. The weight adjustment is done iteratively so that in the steady state, the loudspeaker produces an output that has the same magnitude but opposite phase vis-a-vis the primary noise, thus resulting in ideal noise cancellation. Under such situation, the adaptive filters are seen to model each noise source exactly.

The adaptive filter algorithm analyzed in this paper for multitonal noise cancellation purpose as explained above is the least mean square (LMS) algorithm [6], which is well known for its simplicity, robustness, and excellent numerical error characteristics. Recently, for the special case of two independent noise sources (i.e., $p = 2$), a twin-reference complex LMS algorithm has been analyzed for convergence in [3]. This analysis first considers the 2×2 deterministic input correlation matrix and using its eigenvectors, constructs a product matrix that turns out to be time invariant. The eigenvalues of the product matrix are then evaluated explicitly as functions of the algorithm step size and convergence conditions on the step size are established by restricting the magnitude of each eigenvalue to lie within one. Such an approach, however, cannot be generalized to the case of an arbitrary number of noise sources, since for higher orders, the eigenvalues become intractably complicated functions of the step size and the noise frequencies.

In this paper, we present an alternative approach to the convergence analysis of a multireference complex LMS algorithm that processes p reference signals where p can be any integer. The analysis, like [3], is a deterministic one and to the best of our knowledge, is the first attempt to determine the convergence condition for a general multireference active noise controller. For this, first the eigenstructure of the input correlation matrix is examined to determine its general form. Next, using this, it is shown that for the multireference case too, the $p \times p$ product matrix is a time invariant one, even though the input correlation matrix in this case does not have a unique eigen decomposition unlike the twin-reference problem. The properties of the product matrix are then explored further to show finally that the norm of the weight error vector approaches zero with time, if and only if the step size is chosen from within certain range. An optimal step-size for fastest convergence is then worked out by minimizing the energy of the error sequence which is seen to

Manuscript received February 26, 2003; revised December 3, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Futoshi Asano.

M. Chakraborty is with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India (e-mail :mrityun@ece.iitkgp.ernet.in).

H. Sakai is with the Department of System Science, Graduate School of Informatics, Kyoto University, Kyoto, Japan 606-8501 (e-mail : hsakai@i.kyoto-u.ac.jp).

Digital Object Identifier 10.1109/TSA.2004.840938

be independent of the tonal frequencies and dependent only on the step-size and the number of tonal components p. The claims regarding convergence condition and optimal step size are then verified by simulation studies.

The paper, like [3], however, assumes the secondary path between the loudspeaker and the error microphone to have unity gain and zero phase shift—an ideal assumption that may not always be met exactly in practice, specially when the separation between the loudspeaker and the error microphone is large enough to give rise to an acoustic path with nonnegligible propagation delay and also when the frequency response of the loudspeaker, the power amplifier and the anti-aliasing filter have distorting influence on the tonal components. In such cases, modifications in the LMS algorithm are required to take into account the nonideal nature of the secondary path, resulting in the usage of the so-called filtered-x LMS algorithm ([1], [2], [8]) rather than the standard LMS algorithm. A deterministic convergence analysis for the filtered-x LMS algorithm is, however, much more complicated than the one presented above and has not been addressed in this paper.

The organization of this paper is as follows : Section II presents the multiple reference complex LMS algorithm and Section III provides its convergence analysis in detail. Simulation studies and concluding remarks on the salient aspects of our treatment and observation are given in Section IV. Throughout the paper, matrices and vectors are denoted, respectively, by boldfaced uppercase and lowercase letters, while characters which are not boldfaced denote scalars. Also, $[\cdot]^t$, $[\cdot]^H$, and $[\cdot]^*$ are used to indicate simple transposition, Hermitian transposition and complex conjugation, respectively, and $\|\cdot\|$ denotes the norm of the vector involved.

II. Multiple Reference LMS Algorithm

The filtering processes involved in the multiple reference active noise canceller of Fig. 1 are shown in Fig. 2, where a set of p reference signals $x_i(n) = e^{j\omega_i n}$, $\omega_i \in (0, 2\pi)$, $i = 1, \ldots, p$ are filtered by p complex valued, single tap adaptive filters simultaneously. The reference signal $x_i(n)$, $i = 1, \ldots, p$ is generated from the speed information for the ith rotating machine, usually provided by a tachometer. The frequency ω_i is given by $\omega_i = 2\pi f_i/f_s$, $i = 1, \ldots, p$, where f_s is the sampling frequency adopted and f_i is the analog frequency given in Hertz. The noise generated by, say, the ith rotating machine is modeled as being produced by a filter $H_i(e^{j\omega})$ driven by the single tone $e^{j\omega_i n}$. In other words, the electrical equivalent of the overall noise generated is given by $d(n) = \sum_{i=1}^{p} H_i(e^{j\omega_i})e^{j\omega_i n}$. The output from the p adaptive filters are added to form the signal $y(n)$. Then, assuming the overall transfer function between the loudspeaker and the error microphone, including power amplifiers, analog anti-aliasing filters and sample and hold to be unity [3], $y(n)$ is subtracted from $d(n)$ to generate the error signal $e(n)$ which provides the discrete time, electrical equivalent of the error microphone output. The error signal is used to adapt the filter weights by the so-called complex LMS algorithm [6]. An underlying assumption of the present treatment is that the p frequencies are distinct, i.e., $\omega_i \neq \omega_j$ for $i \neq j$. Otherwise, the

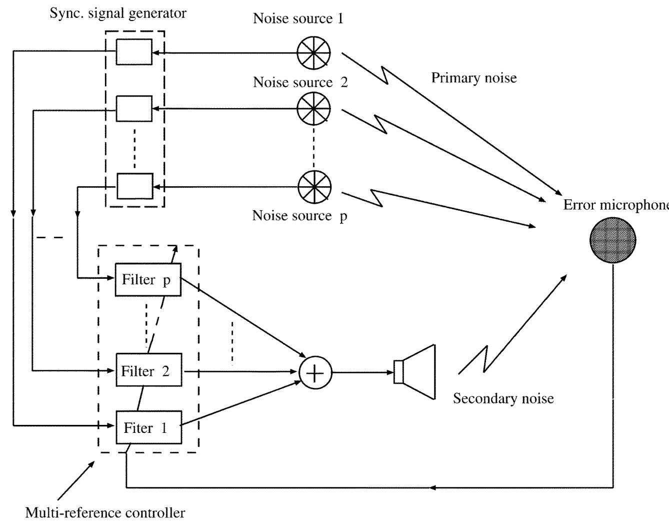

Fig. 1. Physical model of a multireference active noise controller for cancelling multitonal acoustic noise.

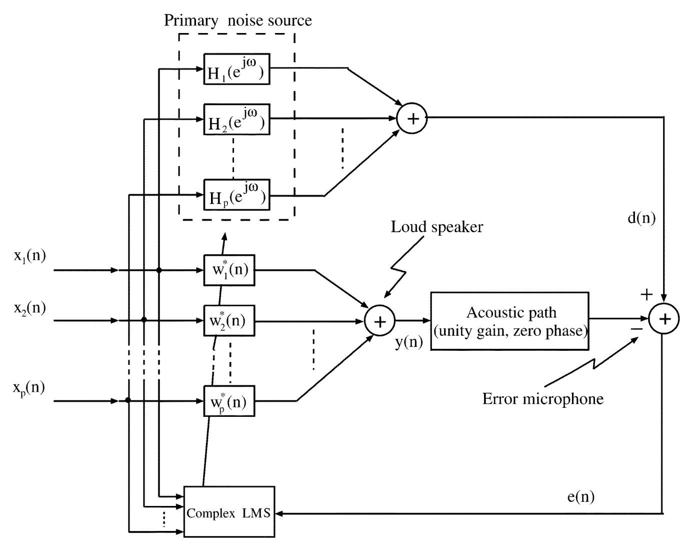

Fig. 2. Adaptive filtering involved in the multireference acoustic noise controller.

problem becomes degenerate since in a situation where r frequencies $(r \leq p)$ are same, no more than $(p - r + 1)$ filters will be necessary to cancel the effect of the noise. Also, similar to [3], we assume without loss of generallity that the phase of each reference signal is zero and amplitude normalized to unity.

Define the input vector $\mathbf{x}(n)$ and the filter weight vector $\mathbf{w}(n)$ at the nth index as $\mathbf{x}(n) = [e^{j\omega_1 n}, e^{j\omega_2 n}, \ldots, e^{j\omega_p n}]^t$ and $\mathbf{w}(n) = [w_1(n), w_2(n), \ldots, w_p(n)]^t$. We then have $y(n) = \mathbf{w}^H(n)\mathbf{x}(n)$ and $e(n) = d(n) - \mathbf{w}^H(n)\mathbf{x}(n)$. The complex LMS algorithm is formulated by first considering the instantaneous squared error function $J(n) = |e(n)|^2$, which, in the current context, is given by

$$J(n) = |d(n)|^2 - \mathbf{w}^H(n)\mathbf{x}(n)d^*(n) - d(n)\mathbf{x}^H(n)\mathbf{w}(n) + \mathbf{w}^H(n)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n). \quad (1)$$

Then, denoting by $\nabla_{\mathbf{w}} J(n)$ the complex gradient vector obtained by differentiating $J(n)$ w.r.t. each tap weight, the weights are updated by following a steepest descent search procedure, as given by

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{\mu}{2}\nabla_{\mathbf{w}} J(n)|_{\mathbf{w}=\mathbf{w}(n)} \tag{2}$$

where μ is the so-called step-size. The gradient $\nabla_{\mathbf{w}} J(n)$ is given by $\nabla_{\mathbf{w}} J(n)|_{\mathbf{w}=\mathbf{w}(n)} = -2\mathbf{x}(n)d^*(n) + 2\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}(n) = -2\mathbf{x}(n)e^*(n)$. Substituting in (2), we obtain the multiple reference complex LMS algorithm as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu\mathbf{x}(n)e^*(n). \tag{3}$$

The step-size μ is to be chosen appropriately for convergence of the above iterative procedure. In Section III, we show that in the context of the active noise controller with multiple, deterministic reference input as described above, a necessary and sufficient condition for the corresponding LMS algorithm to converge is given by : $|1-\mu p| < 1$, or, equivalently, $-1 < 1-\mu p < 1$, implying that μ should be chosen to satisfy : $0 < \mu < 2/p$.

III. Convergence Analysis of the Multiple Reference LMS Algorithm

We define the optimal weight vector $\mathbf{w}_o$ as $\mathbf{w}_o = [H_1^*(e^{j\omega_1}), H_2^*(e^{j\omega_2}), \ldots, H_p^*(e^{j\omega_p})]^t$. The noise signal $d(n)$ is then given as $d(n) = \mathbf{w}_o^H\mathbf{x}(n)$. Ideally, we would want $\mathbf{w}(n)$ to converge to $\mathbf{w}_o$ in some appropriate sense. It is easy to see that if at any index n_0, we have $\mathbf{w}(n_0) = \mathbf{w}_o$, the algorithm converges completely, meaning $e(n) = 0$ and $\mathbf{w}(n) = \mathbf{w}_o$ for all $n \geq n_0$. Conventionally, the input to the LMS algorithm as well as the desired response are random processes and under such case, convergence is reached in mean, i.e., for appropriate choice of the step size, we have $E[\mathbf{w}(n)] \rightarrow \mathbf{w}_o$ as $n \rightarrow \infty$. This, in turn, requires the so-called "independence" assumption requiring statistical independence between $\mathbf{x}(n)$ and $\mathbf{w}(n)$. However, for the active noise controller being considered here, each input is a deterministic signal of a specific type, namely, complex sinusoid of known frequency. Similarly, the desired response $d(n)$ is a deterministic signal, being generated by a set of LTI filters acting on deterministic input. Under such case, it is possible to carry out a deterministic convergence analysis, without requiring the "independence" assumption and by showing that for appropriate choice of μ, $\|\mathbf{v}(n)\|$ actually approaches zero as $n \rightarrow \infty$, where $\mathbf{v}(n)$ is the weight error vector at the nth index and is given by $\mathbf{v}(n) = \mathbf{w}(n) - \mathbf{w}_o$.

Substituting $e^*(n)$ in (3) by $[d^*(n) - \mathbf{x}^H(n)\mathbf{w}(n)]$, then replacing $\mathbf{w}(n)$ by $\mathbf{v}(n)+\mathbf{w}_o$ and noting that $d^*(n) = \mathbf{x}^H(n)\mathbf{w}_o$, it is easy to obtain

$$\mathbf{v}(n+1) = \mathbf{G}(n)\mathbf{v}(n) \tag{4}$$

where

$$\mathbf{G}(n) = \mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^H(n). \tag{5}$$

Applying the recursion in (4) repetitively backward till $n = 0$ and denoting by $\mathbf{v}(0)$ the initial weight error vector, we have

$$\mathbf{v}(n) = \prod_{k=0}^{n-1}\mathbf{G}(k)\mathbf{v}(0). \tag{6}$$

Next, we consider the eigen decomposition of the Hermitian matrix $\mathbf{G}(n)$. Note that the eigen subspaces of $\mathbf{G}(n)$ and $\mathbf{x}(n)\mathbf{x}^H(n)$ are same. Further, the matrix $\mathbf{x}(n)\mathbf{x}^H(n)$ is a rank one matrix with two distinct eigenvalues, one of multiplicity one given by $\|\mathbf{x}(n)\|^2$ and the other of multiplicity $(p-1)$ given by zero. The eigen subspace corresponding to the eigenvalue $\|\mathbf{x}(n)\|^2$ is the range space of $\mathbf{x}(n)\mathbf{x}^H(n)$ and is spanned by $\mathbf{x}(n)$, whereas the eigen subspace corresponding to zero eigenvalue is the null space of $\mathbf{x}(n)\mathbf{x}^H(n)$ and has dimension $(p-1)$. Since the matrix is Hermitian, the null space and the range space are mutually orthogonal. Denoting by $\mathbf{e}_i(n)$, $i = 1, \ldots, p-1$, a set of $(p-1)$ orthonormal vectors spanning the null space of $\mathbf{x}(n)\mathbf{x}^H(n)$ and by $\mathbf{e}_p(n)$ the unit norm vector $(1/\sqrt{p})\mathbf{x}(n)$, we express $\mathbf{G}(n)$ as

$$\mathbf{G}(n) = \mathbf{Q}(n)\mathbf{D}\mathbf{Q}^H(n) \tag{7}$$

where $\mathbf{Q}(n) = [\mathbf{e}_1(n), \ldots, \mathbf{e}_{p-1}(n), \mathbf{e}_p(n)]$ and

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & \ldots & \cdot & 0 \\ 0 & 1 & \ldots & \cdot & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & (1-\mu p) \end{bmatrix}.$$

Substituting (7) in the matrix product in (6), we can write

$$\prod_{k=0}^{n-1}\mathbf{G}(k) = \mathbf{Q}(n-1)\mathbf{D}\mathbf{Q}^H(n-1)\mathbf{Q}(n-2) \cdot\mathbf{D}\mathbf{Q}^H(n-2)\mathbf{Q}(n-3)\ldots\mathbf{Q}(0)\mathbf{D}\mathbf{Q}^H(0). \tag{8}$$

Next we show that though the matrix $\mathbf{Q}(n)$ is dependent on the time index n, the product matrix $\mathbf{B} = \mathbf{D}\mathbf{Q}^H(n)\mathbf{Q}(n-1)$ is time invariant. For this, we first make the following observation.

Theorem 1: Any eigenvector of $\mathbf{G}(n)$ corresponding to eigenvalue 1 is of the form : $[-c_1 e^{j\omega_1 n}, \ldots, -c_{p-1}e^{j\omega_{p-1}n}, (\sum_{i=1}^{p-1} c_i)e^{j\omega_p n}]^t$, $c_i \in \mathcal{C}$.

Proof: The proof of the theorem follows easily by considering a set of $(p-1)$ linearly independent vectors $\tilde{\mathbf{e}}_i(n)$, $i = 1, \ldots, p-1$, where $\tilde{\mathbf{e}}_i(n) = [0, \ldots, 0, -e^{j\omega_i n}, 0, \ldots, 0, e^{j\omega_p n}]^t$ with $(-e^{j\omega_i n})$ constituting the ith element. Clearly, each $\tilde{\mathbf{e}}_i(n)$ is orthogonal to $\mathbf{e}_p(n)$ and thus the set $\tilde{\mathbf{e}}_i(n)$, $i = 1, \ldots, p-1$ forms a basis for the null space of $\mathbf{x}(n)\mathbf{x}^H(n)$. Any vector belonging to the null space of $\mathbf{x}(n)\mathbf{x}^H(n)$, or, equivalently, to the eigen subspace of $\mathbf{G}(n)$ corresponding to eigenvalue 1 is given by a linear combination of the vectors $\tilde{\mathbf{e}}_i(n)$, $i = 1, \ldots, p-1$ and, thus, is of the form : $\left[-c_1 e^{j\omega_1 n}, -c_2 e^{j\omega_2 n}, \ldots, \left(\sum_{i=1}^{p-1} c_i\right) e^{j\omega_p n}\right]^t$. Hence proved. ■

Using theorem 1, we denote $\mathbf{e}_i(n)$, $i = 1, \ldots, p-1$ as : $\mathbf{e}_i(n) = [A_{1i}e^{j\omega_1 n}, \ldots, A_{pi}e^{j\omega_p n}]^t$, with $\sum_{j=1}^{p} A_{ji} = 0$.

Then, the matrix $\mathbf{Q}(n)$ can be written as $\mathbf{Q}(n) = \mathbf{U}(n)\mathbf{A}$, where

$$\mathbf{U}(n) = \begin{bmatrix} e^{j\omega_1 n} & 0 & 0 & \dots & 0 \\ 0 & e^{j\omega_2 n} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & e^{j\omega_p n} \end{bmatrix} \tag{9}$$

and

$$\mathbf{A} = \begin{bmatrix} A_{11} & \dots & A_{1,p-1} & \frac{1}{\sqrt{p}} \\ A_{21} & \dots & A_{2,p-1} & \frac{1}{\sqrt{p}} \\ \cdot & \dots & \cdot & \cdot \\ \cdot & \dots & \cdot & \cdot \\ \cdot & \dots & \cdot & \cdot \\ A_{p1} & \dots & A_{p,p-1} & \frac{1}{\sqrt{p}} \end{bmatrix}. \tag{10}$$

It is then possible to observe the following:

- The matrix $\mathbf{U}(n)$ is unitary, i.e., $\mathbf{U}^H(n)\mathbf{U}(n) = \mathbf{U}(n)\mathbf{U}^H(n) = \mathbf{I}$ for all n.
- Since both $\mathbf{Q}(n)$ and $\mathbf{U}(n)$ are unitary, the matrix $\mathbf{A}$ is also unitary, i.e., $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}$.
- The matrix $\mathbf{B}$ can be written as : $\mathbf{B} = \mathbf{D}\mathbf{Q}^H(n)\mathbf{Q}(n-1) = \mathbf{D}\mathbf{A}^H\mathbf{U}^H(n)\mathbf{U}(n-1)\mathbf{A} = \mathbf{D}\mathbf{A}^H\mathbf{U}(-1)\mathbf{A} = \mathbf{D}\mathbf{C}$, where $\mathbf{C} = \mathbf{A}^H\mathbf{U}(-1)\mathbf{A}$. Clearly, $\mathbf{B}$ does not depend on the time index n.
- The matrix $\mathbf{C}$ is unitary, i.e., $\mathbf{C}^H\mathbf{C} = \mathbf{C}\mathbf{C}^H = \mathbf{I}$.

We now use the above results to prove convergence of the algorithm and to establish convergence condition. For this, we first prove the following.

Theorem 2: For any nonzero $\mathbf{z} \in \mathbf{C}^{p\times 1}$, the last elements (i.e., the pth entries) of the following p vectors : $\mathbf{Cz}, \mathbf{C}^2\mathbf{z}, \dots, \mathbf{C}^p\mathbf{z}$ cannot be zero simultaneously.

Proof: We prove by contradiction. Assume that the pth entry of each of the vectors : $\mathbf{Cz}, \mathbf{C}^2\mathbf{z}, \dots, \mathbf{C}^p\mathbf{z}$ is zero. Now, for $1 \le k \le p$, we have $\mathbf{C}^k\mathbf{z} = \mathbf{A}^H\mathbf{U}^k(-1)\mathbf{Az} = \mathbf{A}^H\mathbf{z}'$, where $\mathbf{z}' = \left[z_1', \dots, z_p'\right]^t = \mathbf{U}^k(-1)\mathbf{Az}$. Also note that each entry of the last row of $\mathbf{A}^H$ is given by $1/\sqrt{p}$. Therefore, if the pth element of $\mathbf{C}^k\mathbf{z}$ is zero, we have $\sum_{i=1}^p z_i' = 0$, or, equivalently

$$\left[e^{-j\omega_1 k}, e^{-j\omega_2 k}, \dots, e^{-j\omega_p k}\right]\mathbf{Az} = 0.$$

Since this is satisfied simultaneously for $k = 1, \dots, p$, we can further write

$$\mathbf{VAz} = \mathbf{0}_{p\times 1} \tag{11}$$

where $\mathbf{0}_{p\times 1}$ is a $p \times 1$ vector of zeros and

$$\mathbf{V} = \begin{bmatrix} e^{-j\omega_1} & \dots & e^{-j\omega_p} \\ e^{-j2\omega_1} & \dots & e^{-j2\omega_p} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ e^{-jp\omega_1} & \dots & e^{-jp\omega_p} \end{bmatrix}.$$

Note that the matrix $\mathbf{V}$ can be written as $\mathbf{V} = \mathbf{V}'\mathbf{U}(-1)$, where

$$\mathbf{V}' = \begin{bmatrix} 1 & \dots & 1 \\ e^{-j\omega_1} & \dots & e^{-j\omega_p} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ e^{-j(p-1)\omega_1} & \dots & e^{-j(p-1)\omega_p} \end{bmatrix}.$$

The matrix $\mathbf{V}'$ is a full rank Vandermonde matrix [7] since, as per our initial assumption, $\omega_1 \ne \omega_2 \ne \dots \ne \omega_p$ with $\omega_i \in (0, 2\pi)$, $i = 1, \dots, p$. From this and also from the fact that both $\mathbf{U}(-1)$ and $\mathbf{A}$ are unitary, it follows that the only solution for (11) is given by $\mathbf{z} = \mathbf{0}_{p\times 1}$, which is a contradiction. Hence proved. ∎

Note : Though the above proof uses the fact that the entries in the last column of $\mathbf{A}$ are all equal ($= 1/\sqrt{p}$), this is, however, not necessary and the only condition required is that no element in the last column of $\mathbf{A}$ should be zero. In fact, (11) can be written in the general case as $\mathbf{V}\tilde{\mathbf{A}}\mathbf{Az} = \mathbf{0}_{p\times 1}$, where $\tilde{\mathbf{A}}$ is a diagonal matrix with ith diagonal entry given by A_{ip}^*, $i = 1, \dots, p$. Clearly, $\tilde{\mathbf{A}}$ is full rank if $A_{ip} \ne 0$, $i = 1, \dots, p$.

We use the above theorem to prove the following important result.

Theorem 3: Given $|1 - \mu p| < 1$ and any nonzero vector $\mathbf{x} \in \mathcal{C}^{p\times 1}$, $||\mathbf{B}^k\mathbf{x}|| < ||\mathbf{x}||$ for $k \ge p$.

Proof: First, we consider the vector $\mathbf{Bx} = \mathbf{DCx}$. Denoting the pth entry of $\mathbf{C}^k\mathbf{x}$ by $[\mathbf{C}^k\mathbf{x}]_p$, it is easily seen that if $[\mathbf{Cx}]_p \ne 0$, then $||\mathbf{Bx}|| = ||\mathbf{DCx}|| < ||\mathbf{Cx}||$ since $|1 - \mu p| < 1$. As $\mathbf{C}$ is unitary, meaning $||\mathbf{Cx}|| = ||\mathbf{x}||$, this implies that $||\mathbf{Bx}|| < ||\mathbf{x}||$. On the other hand, if $[\mathbf{Cx}]_p = 0$, then $\mathbf{Bx} = \mathbf{Cx}$ and thus $||\mathbf{Bx}|| = ||\mathbf{x}||$. Conversely, it also follows that if $||\mathbf{Bx}|| = ||\mathbf{x}||(= ||\mathbf{Cx}||)$, then $[\mathbf{Cx}]_p = 0$. Combining the two cases, we observe $||\mathbf{Bx}|| \le ||\mathbf{x}||$ which can be generalized to include $\mathbf{B}^k\mathbf{x}$. Noting that $||\mathbf{B}^k\mathbf{x}|| = ||\mathbf{B}(\mathbf{B}^{k-1}\mathbf{x})|| \le ||(\mathbf{B}^{k-1}\mathbf{x})||$, we can write

$$\dots \le ||\mathbf{B}^k\mathbf{x}|| \le ||\mathbf{B}^{k-1}\mathbf{x}|| \le \dots \le ||\mathbf{Bx}|| \le ||\mathbf{x}||. \tag{12}$$

We now prove the theorem by contradiction. Since, for $k > p$, $||\mathbf{B}^k\mathbf{x}|| \le ||\mathbf{B}^p\mathbf{x}||$, it is sufficient to prove that $||\mathbf{B}^p\mathbf{x}|| < ||\mathbf{x}||$. Assume that there exists at least one nonzero vector $\mathbf{x} \in \mathcal{C}^{p\times 1}$ so that $||\mathbf{B}^p\mathbf{x}|| = ||\mathbf{x}||$. From (12), it then follows that

$$||\mathbf{B}^p\mathbf{x}|| = ||\mathbf{B}^{p-1}\mathbf{x}|| = \dots = ||\mathbf{Bx}|| = ||\mathbf{x}||. \tag{13}$$

First, we consider $||\mathbf{Bx}|| = ||\mathbf{x}||$. As shown above, this implies that $[\mathbf{Cx}]_p = 0$ and therefore, $\mathbf{Bx} = \mathbf{DCx} = \mathbf{Cx}$. Next, consider $||\mathbf{B}^2\mathbf{x}|| = ||\mathbf{Bx}||$, which can now be written as $||\mathbf{BCx}|| = ||\mathbf{Cx}||$. Using the previous logic, this implies that $[\mathbf{C}^2\mathbf{x}]_p = 0$ and thus, $\mathbf{B}^2\mathbf{x} = \mathbf{DC}^2\mathbf{x} = \mathbf{C}^2\mathbf{x}$. Proceeding like this up to $\mathbf{B}^p\mathbf{x}$, we obtain, $[\mathbf{C}^p\mathbf{x}]_p = [\mathbf{C}^{p-1}\mathbf{x}]_p = \dots = [\mathbf{Cx}]_p = 0$, which contradicts Theorem 2 since $\mathbf{x}$ is nonzero. Hence proved. ∎

We now show that given $\mathbf{x}$ to be any nonzero vector, $\lim_{n\to\infty} ||\mathbf{B}^n\mathbf{x}|| = 0$ if $|1 - \mu p| < 1$. First, if for any finite integer m, $m \ge 1$, $\mathbf{B}^m\mathbf{x}$ is a zero vector under the condition $|1 - \mu p| < 1$, then for all $k > m$ too, $\mathbf{B}^k\mathbf{x}$ is a zero vector and the statement is proved trivially. Such a situation arises when $\mathbf{B}$ is not full rank, meaning $1 - \mu p = 0$ and $\mathbf{x}$ is such a vector that for some integer m, $m \ge 1$, $\mathbf{CB}^{m-1}\mathbf{x}$ lies in the null space of

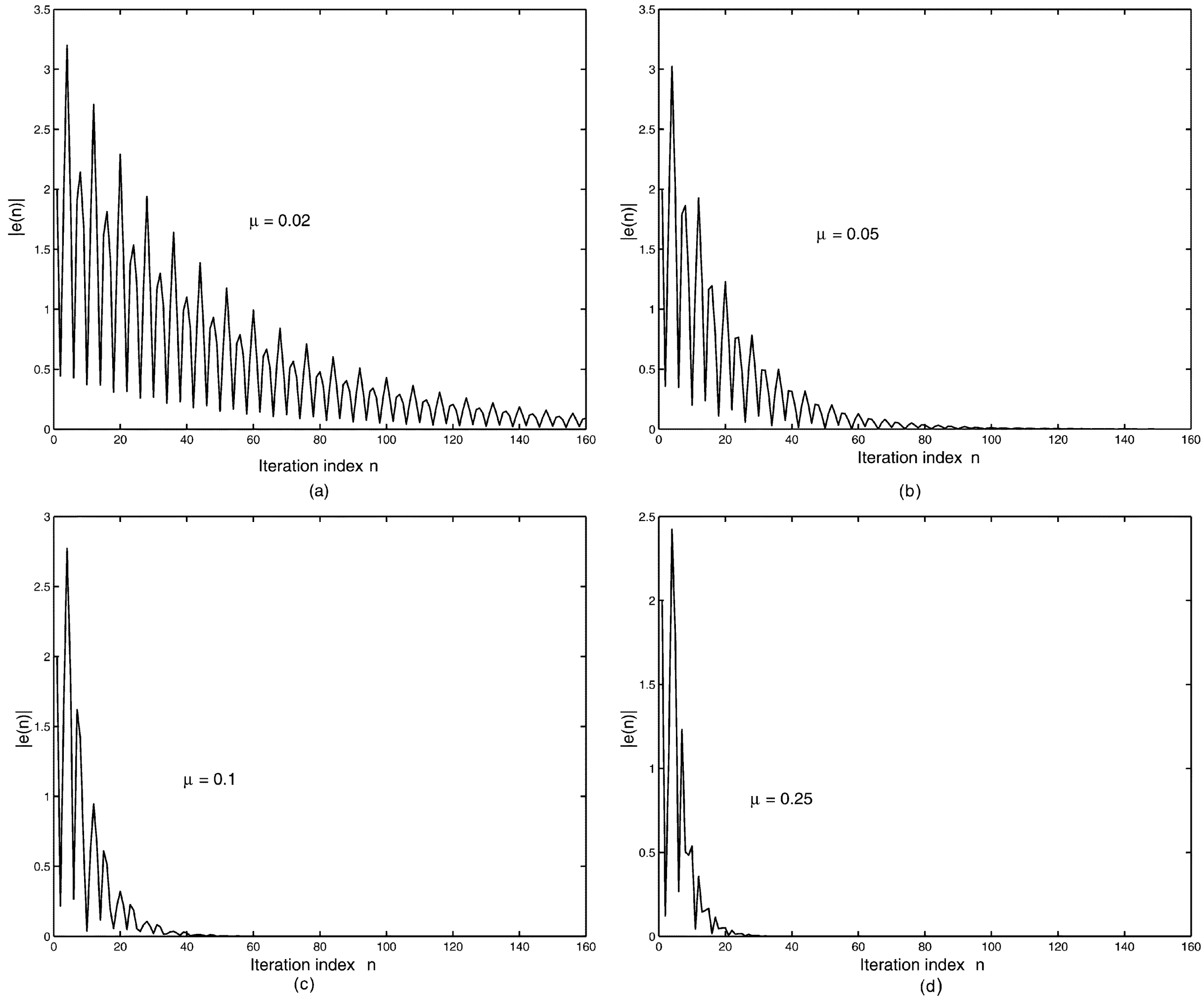

Fig. 3. Learning curves for (a) $\mu = 0.02$. (b) $\mu = 0.05$. (c) $\mu = 0.1$, and (d) $\mu = 0.25$.

$\mathbf{D}$. To prove the convergence in the general case, we discount this possibility and assume that for all integer k, $k \geq 1$, $\mathbf{B}^k\mathbf{x}$ is nonzero, or, equivalently, we assume that either $1-\mu p \neq 0$, or, if $1-\mu p = 0$, then $\mathbf{CB}^{k-1}\mathbf{x}$ does not belong to the null space of $\mathbf{D}$ for any integer $k, k \geq 1$. It then follows from Theorem 3 that $||\mathbf{x}|| > ||\mathbf{B}^p\mathbf{x}|| > ||\mathbf{B}^{2p}\mathbf{x}|| > \cdots$. Since norm of a vector is nonnegative, this along with (12) implies that $\lim_{n\to\infty} ||\mathbf{B}^n\mathbf{x}|| = 0$. Also, it is easy to verify that if $|1-\mu p| > 1$ then, in general, $\cdots \geq ||\mathbf{B}^k\mathbf{x}|| \geq ||\mathbf{B}^{k-1}\mathbf{x}|| \geq \cdots \geq ||\mathbf{Bx}|| \geq ||\mathbf{x}||$ and for $k \geq p$, $||\mathbf{B}^k\mathbf{x}|| > ||\mathbf{x}||$, meaning that in such case, $\lim_{n\to\infty} ||\mathbf{B}^n\mathbf{x}|| = \infty$. [When $|1-\mu p| = 1$, $||\mathbf{B}^n\mathbf{x}|| = ||\mathbf{x}|| > 0$ for all n]. In other words, the condition : $|1-\mu p| < 1$ is both *necessary and sufficient* for convergence of the sequence $||\mathbf{B}^n\mathbf{x}||$ to zero.

Substituting (8) in (6) and using the definition of the matrix $\mathbf{B}$, we then write

$$\mathbf{v}(n) = \mathbf{Q}(n-1)\mathbf{B}^{n-1}\mathbf{DQ}^H(0)\mathbf{v}(0). \tag{14}$$

Since $\mathbf{Q}(n-1)$ is unitary, we have $||\mathbf{v}(n)|| = ||\mathbf{B}^{n-1}\mathbf{z}||$, where $\mathbf{z} = \mathbf{DQ}^H(0)\mathbf{v}(0)$. In the general case, $\mathbf{z}$ is a nonzero vector and therefore, from above, $\lim_{n\to\infty} ||\mathbf{v}(n)|| = 0$ if and only if $|1-\mu p| < 1$, or, equivalently, $0 < \mu < 2/p$. [When $\mathbf{z}$ is a zero vector, $\lim_{n\to\infty} ||\mathbf{v}(n)||$ is zero trivially, which occurs when $\mathbf{D}$ is rank deficient, meaning $1-\mu p = 0$ and $\mathbf{Q}^H(0)\mathbf{v}(0)$ belongs to the null space of $\mathbf{D}$.]

Finally, using the fact that $\lim_{n\to\infty} ||\mathbf{v}(n)||^2 = 0$ for $0 < \mu < 2/p$, we make an interesting observation on $\sum_{k=0}^{\infty} |e(k)|^2$ as given in the theorem below.

Theorem 4: Given $0 < \mu < 2/p$, $\sum_{k=0}^{\infty} |e(k)|^2 = (||\mathbf{v}(0)||^2)/(2\mu - \mu^2 p)$. In other words, $\sum_{k=0}^{\infty} |e(k)|^2$ is independent of the frequencies ω_i, $i = 1, \ldots, p$ and depends only on μ and p.

Proof: First note that the error signal $e(n)$ can be written as $e(n) = d(n) - \mathbf{w}^H(n)\mathbf{x}(n) = d(n) - (\mathbf{w}_o^H + \mathbf{v}^H(n))\mathbf{x}(n) = -\mathbf{v}^H(n)\mathbf{x}(n)$, as $d(n) = \mathbf{w}_o^H\mathbf{x}(n)$. Next, from (4), we write, $||\mathbf{v}(n+1)||^2 = \mathbf{v}^H(n)\mathbf{G}^H(n)\mathbf{G}(n)\mathbf{v}(n)$. Substituting $\mathbf{G}(n)$ by $(\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^H(n))$ as given by (5) and using the above expression of $e(n)$, we can write

$$||\mathbf{v}(n+1)||^2 = ||\mathbf{v}(n)||^2 + |e(n)|^2[\mu^2 p - 2\mu].$$

Applying the recursion on $||\mathbf{v}(n)||^2$ repetitively backward till $n = 0$, we obtain

$$||\mathbf{v}(n+1)||^2 = ||\mathbf{v}(0)||^2 + [\mu^2 p - 2\mu]\sum_{k=0}^{n} |e(k)|^2.$$

By letting n approach infinity on both sides and recalling that for $0 < \mu < 2/p$, $\lim_{n\to\infty} ||\mathbf{v}(n+1)||^2 = 0$, it is then straightforward to write $\sum_{k=0}^{\infty} |e(k)|^2 = (||\mathbf{v}(0)||^2)/(2\mu - \mu^2 p)$. Hence proved. ■

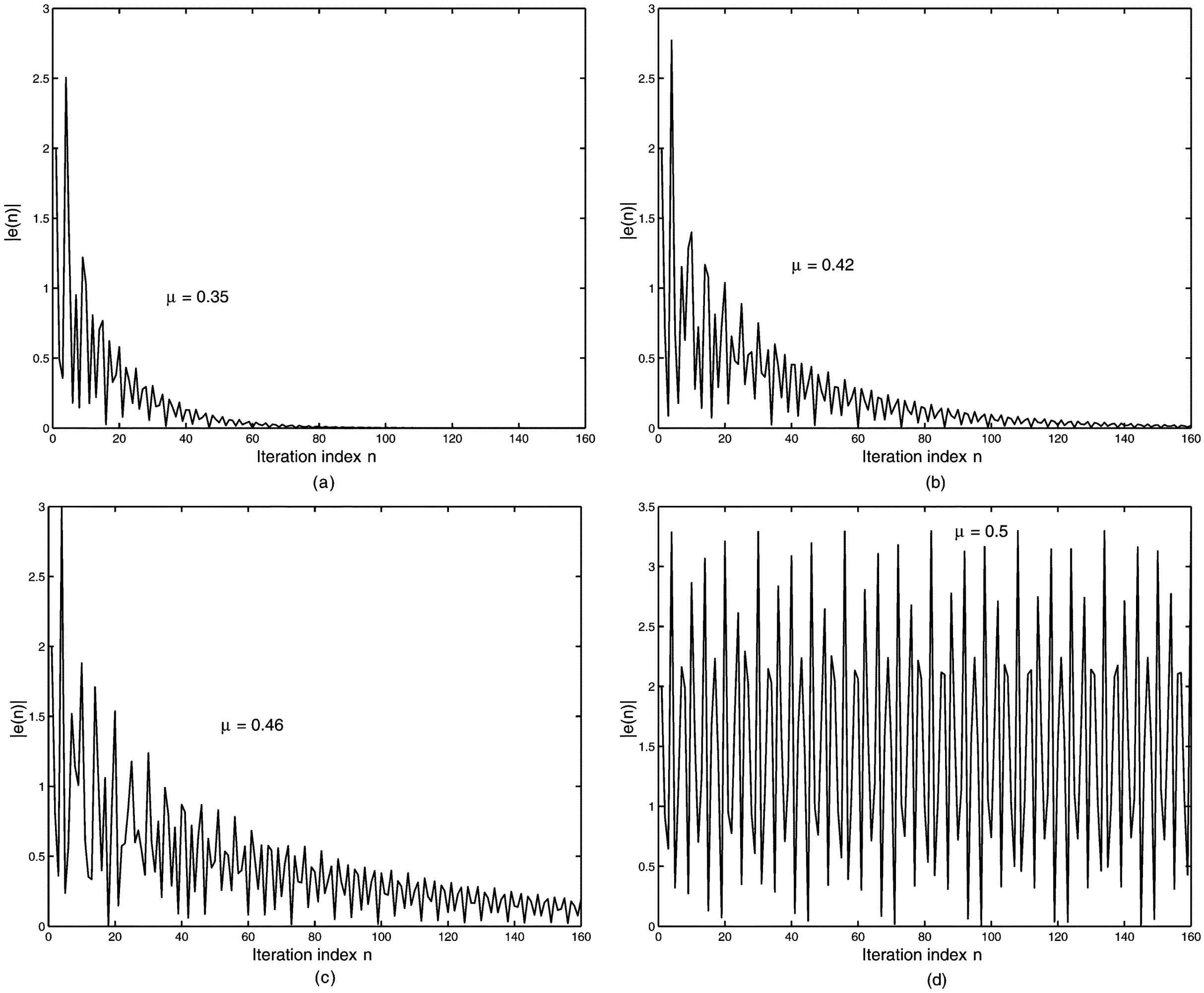

Fig. 4. Learning curves for (a) $\mu = 0.35$. (b) $\mu = 0.42$. (c) $\mu = 0.46$. (d) $\mu = 0.5$.

The above theorem can be used to determine the optimal step size for fastest convergence as discussed in the following section.

IV. Simulation Studies, Discussion, and Conclusions

In this paper, we have considered an active noise controller for cancelling tonal acoustic noise and determined the convergence condition for the multireference complex LMS algorithm needed to train the controller with no restriction on the number of reference signals. Interestingly, the convergence condition : $0 < \mu < 2/p$ derived for unit power complex sinusoid input is identical to the stochastic case where p unit variance random signals are filtered by a set of p single tap LMS based adaptive filters to estimate a desired random process. However, the deterministic convergence analysis ensures that in the absence of modeling error and any observation noise, the magnitude of the error signal actually approaches zero, while, in the stochastic case, the filter coefficients converge to the optimal weights only in the mean and thus a residual mean square error (mse) is observed in the steady state mse. It is also interesting to observe that a necessary and sufficient condition for $\lim_{n\to\infty} ||\mathbf{B}^n\mathbf{z}|| = 0$ for any $\mathbf{z}$ is that the spectral radius of $\mathbf{B}$, denoted by $\rho(\mathbf{B})$ and given by $|\lambda_{\max}|$ ($\lambda_{\max}$: eigenvalue of $\mathbf{B}$ with highest magnitude) should satisfy $\rho(\mathbf{B}) < 1$ [7]. Comparing the two convergence conditions, both being necessary as well as sufficient, we can then infer that $|\lambda_i| < 1$, iff $0 < \mu < 2/p$, where λ_i denotes the ith eigenvalue of $\mathbf{B}$, $i = 1, \ldots, p$.

The reference signals for the LMS algorithm are assumed to be of the form of unit magnitude complex sinusoid, whereas in real life, a tonal signal is either a cosinusoidal or a sinusoidal function of the time index n. However, the tonal component $A\cos(\omega n+\phi)$ generated by one rotating machine can be modeled as being generated by two systems, having input $e^{j\omega n}$ and $e^{-j\omega n}$ and gain of $(A/2)e^{j\phi}$ and $(A/2)e^{-j\phi}$ respectively at frequency ω. In our simulation studies, we considered two tonal noise frequencies of 125 Hz and 250 Hz and a sampling frequency of 1 kHz which gives rise to $\omega_1 = \pi/4$ and $\omega_2 = \pi/2$. The discrete time versions of the tonal acoustic noise considered were of the form : $2\sqrt{2}\cos(\pi n/4+\pi/4)$ and $0.5\sin(\pi n/2)$. As per the formulation adopted in this paper, we then have $p = 4$ with $\mathbf{x}(n) = [e^{j\pi n/4}, e^{-j\pi n/4}, e^{j\pi n/2}, e^{-j\pi n/2}]^t$ and the optimal weight vector $\mathbf{w}_o = [1+j, 1-j, -j/4, +j/4]^H$. The total tonal noise generated is given by $\mathbf{w}_o^H\mathbf{x}(n)$ and zero mean, additive white Gaussian noise of variance $10^{-8}(-80\,\mathrm{dB})$ is added to it to constitute $d(n)$. The multireference LMS algorithm is run for increasing values of μ in the range : $0 < \mu < 2/p = 0.5$ and $|e(n)|$ is plotted vis-a-vis n to obtain the learning curves, as shown in Fig. 3 and 4. Additionally, we also plot $|e(n)|$ in dB (i.e., $20\log_{10}(|e(n)|)$) against n

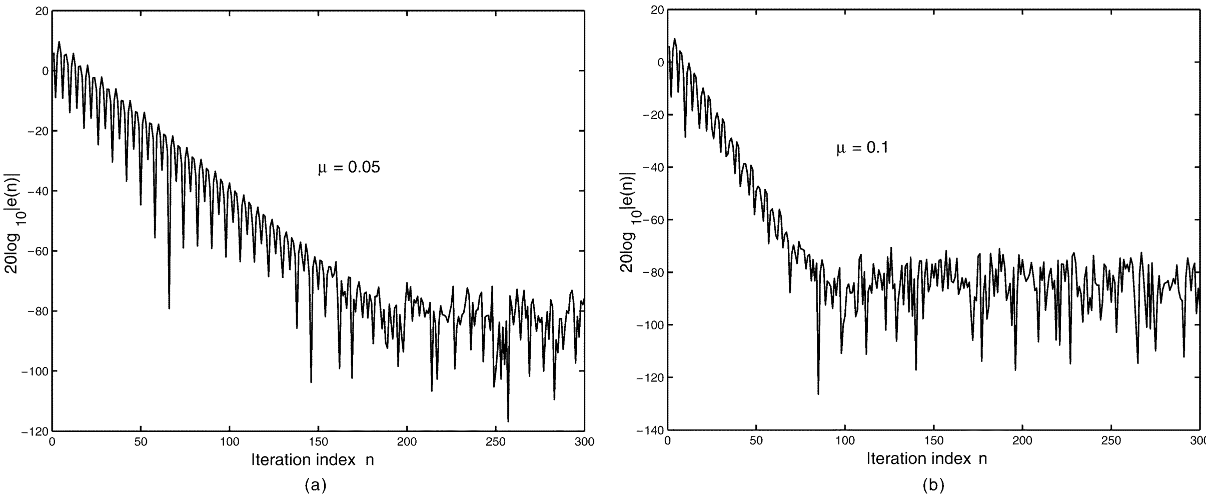

Fig. 5. Learning curves in dB for (a) $\mu = 0.05$. (b) $\mu = 0.1$.

in Fig. 5 for $\mu = 0.05$ and $\mu = 0.1$ to show the presence of residual error in the steady state contributed by the additive white Gaussian noise.

An inspection of Fig. 3 and 4 reveals that there is certain range of values of μ that gives rise to fastest convergence and as μ deviates further and further from this range, the rate of convergence deteriorates. Thus, in Fig. 3, it is seen that the convergence speed improves as μ progressively takes the following values : 0.02, 0.05, 0.1 and 0.25. On the other hand, Fig. 4 shows that as μ increases further and takes the values : 0.35, 0.42 and 0.46, the rate of convergence decreases and for $\mu = 0.5$, the algorithm does not converge. An explanation for this phenomenon may be provided by recalling, from Theorem 4, that the energy of the error sequence $e(n)$, under convergence condition, is given by $(\|\mathbf{v}(0)\|^2)/(2\mu - \mu^2 p)$. It is easy to verify that the function $(\|\mathbf{v}(0)\|^2)/(2\mu - \mu^2 p)$ has a unique minima at $\mu = 1/p$. Since, every converging $e(n)$ is a decreasing function of n, the minimum energy error sequence, for all practical purposes, will have speed of convergence higher than that of other error sequences and thus $\mu = 1/p$ provides the optimal step size for fastest convergence. This explains how, in our simulation studies, fastest convergence takes place at $\mu = 1/p = 0.25$ as shown in Fig. 3(d) and how the convergence becomes slower as μ deviates from this optimal value.

References

[1] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems*. New York: Wiley, 1996.
[2] C. R. Fuller, S. J. Elliot, and P. A. Nelson, *Active Control of Vibrations*. New York: Academic, 1996.
[3] S. Johansson, S. Nordebo, and I. Claesson, "Convergence analysis of a twin-reference complex least-mean-squares algorithm," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 3, pp. 213–221, May 2002.
[4] S. Johansson, I Claesson, S. Nordebo, and P. Sjösten, "Evaluation of multiple reference active noice control algorithms on Dornier 328 aircraft data," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 4, pp. 473–477, Jul. 1999.
[5] S. J. Elliot, P. A. Nelson, I. M. Stothers, and C. C. Boucher, "In-flight experiments on the active control of propeller-induced cabin noise," *J. Sound Vibr.*, vol. 140, no. 2, pp. 219–238, 1990.
[6] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, MA: Univ. Press, 1985.
[8] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

Mrityunjoy Chakraborty (M'94–SM'99) received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta, India, in 1983 and the M.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology, Kanpur and Delhi, India, in 1985 and 1994, respectively.

He joined the Indian Institute of Technology, Kharagpur, as a Lecturer in 1994, and is now a Professor of electronics and electrical communications engineering. He visited the National University of Singapore from 1998 to 1999 and Kyoto University, Kyoto, Japan, in 2002 and 2003 under a Kyoto University foundation fellowship and a JSPS fellowship. His research interests include digital and adaptive signal processing, VLSI architectures for signal processing, and signal processing applications in wireless communications.

Hideaki Sakai (M'78–SM'02) received the B.E. and Dr. Eng. degrees in applied mathematics and physics from Kyoto University, Kyoto, Japan, in 1972 and 1981, respectively.

From 1975 to 1978, he was with Tokushima University, Tokushima, Japan. He is currently a Professor in the Department of Systems Science, Graduate School of lnformatics, Kyoto University, Kyoto, Japan. He spent six months from 1987 to 1988 at Stanford University, Stanford, CA, as a Visiting Scholar. His research interests are in the area of adaptive signal processing and time series analysis. He has been an Associate Editor of the *IEICE Trans. Fundamentals*, from 1996 to 2000 and is on the editorial board of *EURASIP Journal of Applied Signal Processing.*

Dr. Sakai was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2001.