Exact Convergence Analysis of Adaptive Filter Algorithms Without the Persistently Exciting Condition

Hideaki Sakai, *Fellow, IEEE*, Jun-Mei Yang, *Student Member, IEEE*, and Tetsuya Oka

***Abstract*—Exact convergence analysis of the recursive least square and least mean square (LMS) algorithms in adaptive filtering is presented for the case of sinusoidal signal cancellation without the persistently exciting condition. This situation occurs when the number of tap coefficients of the adaptive filter exceeds that of the complex sinusoids in the input signal. The convergent point of both algorithms is shown to be the one determined by the pseudo inverse of the deterministic covariance matrix. The convergence proof for the LMS algorithm is based on the Lyapunov function method. Finally, the validity of the obtained results is supported by simulation results.**

***Index Terms*—Adaptive filter algorithms, exact convergence analysis, persistently exciting condition, sinusoidal noise cancellation.**

I. Introduction

ADAPTIVE filters are used in many areas of signal processing applications such as noise cancellation, channel equalization, and system identification. For adaptation the recursive least squares (RLS) and least mean square (LMS) algorithms are often used [1].

For tonal noise generated by periodic systems such as rotating machines, the anti-noise generated by a set of adaptive weights is used to cancel the noise. A method of eliminating the multiple sinusoidals or other periodic interference by using a tapped delay line adaptive filter was proposed in [2]. Applying the z-transform to the adaptive algorithm, it was shown that under certain conditions the system can be approximated by a linear time-invariant system whose transfer function has notch characteristic. In [3] and [4], rigorous results about convergence of the LMS algorithm have been derived where complex sinusoids with unit amplitudes and known frequencies are used as the input signals.

In this paper, a similar narrowband noise canceling problem with the one in [2] is considered where the input (reference) signal $x(n)$ to the adaptive filter is a sum of K complex sinusoids $A_k e^{j\omega_k n}$ with *unknown* non-zero amplitudes A_k and *unknown* distinct frequencies ω_k $(k = 1, \ldots, K)$. For real signals, frequencies ω_k and $-\omega_k$ are paired with the complex conjugate amplitude A_k and A_k^*, where $(\cdot)^*$ denotes the complex conjugate. The desired signal $d(n)$ is formed by passing $x(n)$ through a fixed but unknown plant with a stable transfer function $P(z)$ which may be of finite-impulse-response (FIR) or infinite-impulse-response (IIR) type. This looks like a typical system identification problem but our main aim is not to identify the parameters of $P(z)$ consistently but to cancel $d(n)$ by $y(n)$, the output of the adaptive filter. The FIR adaptive filter has L tap coefficients $w_i(n)$ $(i = 0, \ldots, L-1)$ which are adjusted so as to let the error signal $e(n)$ approach zero by the RLS or LMS algorithm.

If $L \leq K$, the persistently exciting (PE) condition is met and by the least squares (LS) identification algorithm after L time steps $w_i(n)$ becomes the corresponding tap of $P(z)$, if it has an $(L-1)$th-order FIR structure [5]. But if $L > K$, then the deterministic covariance matrix is rank deficient, that is, the PE condition is not satisfied and the corresponding normal equation becomes indeterminate (see Appendix I for explanation of the PE condition). There seems to be no rigorous treatment for this case in the literature. This case is important in practical applications, since often we do not know K in advance and a larger L should be used for perfect noise cancellation as will be discussed later.

In this paper, first, we show that the RLS algorithms with and without the forgetting factor converge to a unique point which is the minimum norm solution of the above normal equation with zero error signal. Second, we show that the LMS algorithm converges to the same point and an exact upper limit of the step size is derived as the stability condition for the LMS adaptive filter. This limit has not been rigorously discussed in the literature. This result is used to design an LMS adaptive filter whose performance depends on the selection of a step size parameter. Since the LMS algorithm for this case is expressed as a time-varying linear system, its rigorous convergence analysis is much more complicated than that for RLS. We use the Lyapunov function method in [4] with the same argument as in [3]. Finally, to justify the theoretical assertions some numerical simulation results are demonstrated.

II. Conditions of Perfect Noise Cancellation

The input signal $x(n)$ is passed through the adaptive filter in Fig. 1, where the input vector and the tap weight vector are defined by

$$\begin{aligned} \boldsymbol{x}(n) &= [x(n)\, x(n-1)\, \ldots\, x(n-L+1)]^T \\ \boldsymbol{w}(n) &= [w_0(n)\, w_1(n)\, \ldots\, w_{L-1}(n)]^T \end{aligned} \tag{1}$$

Manuscript received December 25, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jonathon A. Chambers.

The authors are with the Department of Systems Science, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: hsakai@i.kyoto-u.ac.jp; jmyang@sys.i.kyoto-u.ac.jp; oka@sys.i.kyoto-u.ac.jp).

Digital Object Identifier 10.1109/TSP.2006.890930

1053-587X/$25.00 © 2007 IEEE

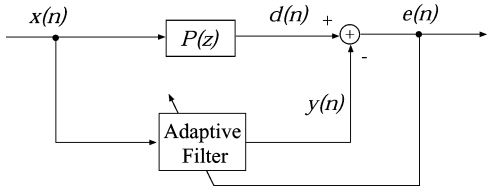


Fig. 1. Block diagram of the adaptive noise canceling.

where $(\cdot)^T$ denotes the transpose and a delay is implicitly included in the adaptive algorithm. The input signal $x(n)$ is assumed to be a sum of complex sinusoids so that we have

$$x(n)=\sum_{k=1}^{K}A_k e^{j\omega_k n},\quad \boldsymbol{x}(n)=\sum_{k=1}^{K}A_k\boldsymbol{\phi}_k e^{j\omega_k n} \tag{2}$$

with

$$\boldsymbol{\phi}_k=\left[1\ e^{-j\omega_k}\ \cdots\ e^{-j(L-1)\omega_k}\right]^T. \tag{3}$$

The output $y(n)$ of the adaptive filter is

$$y(n)=\boldsymbol{w}^H(n)\boldsymbol{x}(n)=\sum_{k=1}^{K}A_k\boldsymbol{w}^H(n)\boldsymbol{\phi}_k e^{j\omega_k n} \tag{4}$$

where $(\cdot)^H$ denotes the Hermitian transpose. Since $P(z)$ is stable, $P(z)$ is expanded as

$$P(z)=\sum_{i=0}^{\infty}p_i z^{-i}$$

so that the steady-state response to $A_k e^{j\omega_k n}$ is

$$\sum_{i=0}^{\infty}p_i A_k e^{j\omega_k(n-i)}=A_k\left(\sum_{i=0}^{\infty}p_i e^{-j\omega_k i}\right)e^{j\omega_k n}.$$

Hence, the desired signal $d(n)$ is expressed by

$$d(n)=\sum_{k=1}^{K}A_k P_k e^{j\omega_k n},\quad P_k=P(e^{j\omega_k}) \tag{5}$$

where $P(e^{j\omega})$ is the frequency response of $P(z)$ at $z=e^{j\omega}$.

In the steady state $\boldsymbol{w}(n)=\boldsymbol{w}$, from (4) and (5) perfect noise cancellation is possible if

$$\sum_{k=1}^{K}A_k\boldsymbol{w}^H\boldsymbol{\phi}_k e^{j\omega_k n}=\sum_{k=1}^{K}A_k P_k e^{j\omega_k n} \tag{6}$$

so that $\boldsymbol{w}$ must satisfy

$$\boldsymbol{\Phi}\boldsymbol{w}=\boldsymbol{\psi} \tag{7}$$

where

$$\boldsymbol{\Phi}^H=[\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \cdots\ \boldsymbol{\phi}_K],\ \boldsymbol{\psi}=[P_1\ P_2\ \cdots\ P_K]^H \tag{8}$$

and $\boldsymbol{\phi}_k$ $(k=1,\ldots,K)$ are linearly independent for $L\geq K$. If $L<K$, (7) is an over determined system. In general, $\boldsymbol{\psi}$ does not lie in the range space of $\boldsymbol{\Phi}$, so there is no solution and perfect noise cancellation is not attained. One generic exception is the case where $P(z)$ is of $(L-1)$th-order FIR, that is, $P_k=\boldsymbol{w}_p^H\boldsymbol{\phi}_k$ with the impulse response $\boldsymbol{w}_p=(w_{p,0}w_{p,1},\ldots,w_{p,L-1})^T$. For this case (7) has a unique solution $\boldsymbol{w}_p$. If $L=K$, obviously (7) has a unique solution. If $L>K$, (7) is an under-determined system whose solution is not unique. But $\boldsymbol{\Phi}$ is of full row rank, so using the (Moore–Penrose) pseudo inverse $\boldsymbol{\Phi}^+$ of $\boldsymbol{\Phi}$, one solution is

$$\boldsymbol{w}_{\rm opt}=\boldsymbol{\Phi}^+\boldsymbol{\psi}=\boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\boldsymbol{\psi} \tag{9}$$

which is known as the minimum norm solution. In summary, for the case $L\geq K$, perfect noise cancellation is possible for $P(z)$ of any stable plant, if the weight vector is set to (9). In the following, we will show that both of the two typical RLS and LMS adaptive algorithms converge to $\boldsymbol{w}_{\rm opt}$ in (9) under some condition for LMS.

III. Analysis of RLS Algorithm

In this section, we present an analysis of the RLS algorithm for the case $L>K$ with and without the forgetting factor. It is well-known that the LS estimate at time n is given by the normal equation

$$\boldsymbol{Q}(n)\boldsymbol{w}(n)=\boldsymbol{q}(n) \tag{10}$$

where the deterministic covariance matrix $\boldsymbol{Q}(n)$ and cross covariance vector $\boldsymbol{q}(n)$ are defined by

$$\begin{aligned}\boldsymbol{Q}(n)&=\sum_{i=1}^{n}\lambda^{n-i}\boldsymbol{x}(i)\boldsymbol{x}^H(i)\\ \boldsymbol{q}(n)&=\sum_{i=1}^{n}\lambda^{n-i}\boldsymbol{x}(i)d^*(i)\end{aligned} \tag{11}$$

for $n>0$, where for simplicity, we assume that $x(-L+1),\ldots,x(0)$ are available and the forgetting factor λ is used with $0<\lambda\leq 1$, so that from (2), (5), and (11)

$$\boldsymbol{Q}(n)=\sum_{k=1}^{K}\sum_{l=1}^{K}A_k A_l^*\boldsymbol{\phi}_k\boldsymbol{\phi}_l^H D_n(\omega_k-\omega_l) \tag{12}$$

$$\boldsymbol{q}(n)=\sum_{k=1}^{K}\sum_{l=1}^{K}A_k A_l^* P_l^*\boldsymbol{\phi}_k D_n(\omega_k-\omega_l) \tag{13}$$

where

$$D_n(\omega)=\sum_{i=1}^{n}\lambda^{n-i}e^{j\omega i}. \tag{14}$$

Defining the matrix $\boldsymbol{D}(n)$ whose (k,l)th element is $D_n(\omega_k-\omega_l)$, (12) and (13) can be expressed as

$$\boldsymbol{Q}(n)=(\boldsymbol{A}\boldsymbol{\Phi})^H\boldsymbol{D}(n)\boldsymbol{A}\boldsymbol{\Phi},\quad \boldsymbol{q}(n)=(\boldsymbol{A}\boldsymbol{\Phi})^H\boldsymbol{D}(n)\boldsymbol{A}\boldsymbol{\psi} \tag{15}$$

where

$$\boldsymbol{A}\triangleq\operatorname{diag}(A_1^*,A_2^*,\ldots,A_K^*). \tag{16}$$

From (14) the following quadratic form:

$$\boldsymbol{\eta}^H \boldsymbol{D}(n)\boldsymbol{\eta} = \sum_{i=1}^{n} \lambda^{n-i} \left| \sum_{k=1}^{K} \eta_k e^{-j\omega_k i} \right|^2$$

with $\boldsymbol{\eta} = [\eta_1 \eta_2, \ldots, \eta_K]^T$ is nonnegative and for $n \geq K$ becomes zero only for $\boldsymbol{\eta} = 0$, since

$$\sum_{k=1}^{K} \eta_k e^{-j\omega_k i} = 0 \quad (i = 1, \ldots, K)$$

means $\boldsymbol{\eta} = 0$ for distinct frequencies $\omega_1, \ldots, \omega_K$. Hence, $\boldsymbol{D}(n)$ is positive definite for $n \geq K$ and is expressed as

$$\boldsymbol{D}(n) = \boldsymbol{G}^H(n)\boldsymbol{G}(n) \tag{17}$$

where $\boldsymbol{G}(n)$ is a $K \times K$ nonsingular matrix. Thus, from (15) and (17) we have

$$\boldsymbol{Q}(n) = (\boldsymbol{R}(n)\boldsymbol{\Phi})^H \boldsymbol{R}(n)\boldsymbol{\Phi}, \quad \boldsymbol{q}(n) = (\boldsymbol{R}(n)\boldsymbol{\Phi})^H \boldsymbol{R}(n)\boldsymbol{\psi} \tag{18}$$

where $\boldsymbol{R}(n) = \boldsymbol{G}(n)\boldsymbol{A}$ is nonsingular. From (18), we note that $\boldsymbol{Q}(n)$ is rank deficient and (10) is indeterminate, that is, the PE condition is not satisfied. So it is unclear to which point the LS estimate is approaching.

Now, we consider the conventional RLS algorithm

$$\begin{aligned} \boldsymbol{w}(n) &= \boldsymbol{w}(n-1) + \tilde{\boldsymbol{Q}}^{-1}(n)\boldsymbol{x}(n)e^*(n) \\ e(n) &= d(n) - \boldsymbol{w}^H(n-1)\boldsymbol{x}(n) \end{aligned} \tag{19}$$

with

$$\tilde{\boldsymbol{Q}}(n) = \lambda \tilde{\boldsymbol{Q}}(n-1) + \boldsymbol{x}(n)\boldsymbol{x}^H(n) \tag{20}$$

and the initialization

$$\boldsymbol{w}(0) = \boldsymbol{0}, \quad \tilde{\boldsymbol{Q}}(0) = \xi \boldsymbol{I} \quad (\xi > 0) \tag{21}$$

where $\boldsymbol{I}$ is the identity matrix and $\tilde{\boldsymbol{Q}}^{-1}(n)$ is updated by using the matrix inversion lemma [1]. From (11), (20), and (21), we note that

$$\tilde{\boldsymbol{Q}}(n) = \lambda^n \xi \boldsymbol{I} + \boldsymbol{Q}(n). \tag{22}$$

First, we treat the case for $\lambda < 1$. Since $\tilde{\boldsymbol{Q}}(n)$ is nonsingular, in the RLS algorithm, (10) is replaced by

$$\boldsymbol{w}(n) = \tilde{\boldsymbol{Q}}^{-1}(n)\boldsymbol{q}(n). \tag{23}$$

Let the eigenspace decomposition of $\boldsymbol{Q}(n)$ be

$$\boldsymbol{Q}(n) = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H \tag{24}$$

where for notational simplicity, we omit the dependence on n in the RHS of (24) and $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K \boldsymbol{u}_{K+1}, \ldots, \boldsymbol{u}_L]$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_K, 0, \ldots, 0)$ with $\boldsymbol{u}_i$ eigenvectors and λ_i, the corresponding eigenvalues, respectively. From (12) and (13), it is obvious that $\boldsymbol{u}_{K+1}, \ldots, \boldsymbol{u}_L$ corresponding to zero eigenvalues are orthogonal to $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K$ and, hence, $\boldsymbol{q}(n)$. So, from (23) and (24), we have

$$\begin{aligned} \boldsymbol{w}(n) &= \boldsymbol{U}(\lambda^n \xi \boldsymbol{I} + \boldsymbol{\Lambda})^{-1} \mathrm{diag}(\overbrace{1, \ldots, 1}^{K}, 0, \ldots, 0)\boldsymbol{U}^H \boldsymbol{q}(n) \\ &= \boldsymbol{U}\mathrm{diag}\left(\frac{1}{\lambda_1 + \lambda^n \xi}, \ldots, \frac{1}{\lambda_K + \lambda^n \xi}, \, 0, \ldots, 0 \right) \boldsymbol{U}^H \boldsymbol{q}(n). \end{aligned} \tag{25}$$

As $n \to \infty$, the matrix in the RHS of (25) tends to $\boldsymbol{U}\mathrm{diag}(1/\lambda_1, \ldots, 1/\lambda_K, 0, \ldots, 0)\boldsymbol{U}^H = \boldsymbol{Q}^+(n)$. This argument is adopted from [6, Th. 3.4] within our RLS case. Hence, $\boldsymbol{w}(n) \to \boldsymbol{Q}^+(n)\boldsymbol{q}(n)$ as $n \to \infty$. Using the identity $\boldsymbol{B}^+ = (\boldsymbol{B}^H \boldsymbol{B})^+ \boldsymbol{B}^H$ for general matrix $\boldsymbol{B}$ in [6, Th. 3.8] and noting that $\boldsymbol{R}(n)\boldsymbol{\Phi}$ in (18) is of full row rank, we have

$$\begin{aligned} \boldsymbol{Q}^+(n)\boldsymbol{q}(n) &= \left((\boldsymbol{R}(n)\boldsymbol{\Phi})^H \boldsymbol{R}(n)\boldsymbol{\Phi} \right)^+ (\boldsymbol{R}(n)\boldsymbol{\Phi})^H \boldsymbol{R}(n)\boldsymbol{\psi} \\ &= (\boldsymbol{R}(n)\boldsymbol{\Phi})^+ \boldsymbol{R}(n)\boldsymbol{\psi} \\ &= (\boldsymbol{R}(n)\boldsymbol{\Phi})^H \left(\boldsymbol{R}(n)\boldsymbol{\Phi}\boldsymbol{\Phi}^H \boldsymbol{R}^H(n) \right)^{-1} \boldsymbol{R}(n)\boldsymbol{\psi} \\ &= \boldsymbol{w}_{\mathrm{opt}}. \end{aligned} \tag{26}$$

That is, $\boldsymbol{w}(n)$ converges to $\boldsymbol{w}_{\mathrm{opt}}$ in (9).

Next, we consider the case $\lambda = 1$. In this case, the diagonal elements of $\boldsymbol{D}(n)$ are n so that we replace (22) by $\tilde{\boldsymbol{Q}}(n)/n = \xi \boldsymbol{I}/n + \boldsymbol{Q}(n)/n$ and (23) by $\boldsymbol{w}(n) = (\tilde{\boldsymbol{Q}}(n)/n)^{-1}(\boldsymbol{q}(n)/n)$, respectively. Applying the same argument after (24) to $\boldsymbol{Q}(n)/n$, we see that $(\tilde{\boldsymbol{Q}}(n)/n)^{-1}(\boldsymbol{q}(n)/n) \to (\boldsymbol{Q}(n)/n)^+(\boldsymbol{q}(n)/n)$ and from (18) we conclude that $\boldsymbol{w}(n)$ converge to $\boldsymbol{w}_{\mathrm{opt}}$ in this case, too. If the observation noise $v(n)$ is added to $d(n)$, its effect does not change the convergence property. This can be seen as follows. From

$$\frac{1}{n}\boldsymbol{q}(n) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}(i)d^*(i) + \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}(i)v^*(i) \tag{27}$$

$\boldsymbol{q}(n)/n$ is still in the space spanned by $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_K$ and the covariance matrix of the second term in (27) is given by $\sigma^2 \sum_{i=1}^{n} \boldsymbol{x}(i)\boldsymbol{x}^H(i)/n^2$ for white noise $v(n)$ with zero mean and variance σ^2. Since $\|\boldsymbol{x}(n)\| \triangleq (\boldsymbol{x}^H(n)\boldsymbol{x}(n))^{1/2}$ is bounded by some constant, this covariance matrix converges to zero and the effect of the observation noise fades out. But, of course, taking λ to 1 means a loss of tracking ability for nonstationary situations such as time variation of signal characteristics.

IV. Analysis of LMS Algorithm

In this section, we give a rigorous convergence analysis of the LMS algorithm for the case $L > K$. Though it is simpler than the RLS algorithm, the analysis is much harder. It is given by

$$\begin{aligned} \boldsymbol{w}(n+1) &= \boldsymbol{w}(n) + \mu \boldsymbol{x}(n)e^*(n) \\ e(n) &= d(n) - \boldsymbol{w}^H(n)\boldsymbol{x}(n) \end{aligned} \tag{28}$$

where μ is a positive step size [7], [8]. Here, we assume that $x(-L+1), \ldots, x(0)$ are available. The region guaranteeing the

convergence will be discussed later. If the initial vector $\boldsymbol{w}(0)$ is taken to $\mathbf{0}$, from (28) $\boldsymbol{w}(n)$ is written as

$$\boldsymbol{w}(n) = \sum_{i=1}^{K} h_i(n)\boldsymbol{\phi}_i. \tag{29}$$

This may be regarded as a kind of discrete Fourier expansion though $\boldsymbol{\phi}_i$'s are not orthogonal. From (2), (5), and (29), we have

$$h_i(n+1) = h_i(n) - \mu \sum_{k=1}^{K} A_i A_k^* e^{j(\omega_i-\omega_k)n} \sum_{l=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l(n) + \mu \sum_{k=1}^{K} A_i A_k^* P_k^* e^{j(\omega_i-\omega_k)n} \quad (i = 1, \ldots, K). \tag{30}$$

The derivation of this equation is given in Appendix II. This is a time-varying multivariable linear system with sinusoidal driving signals. If $\boldsymbol{w}(0)$ is any vector, the component in the range space of $\bar{\boldsymbol{\Phi}}^H$ can be absorbed in (29). The component in the orthogonal compliment of $\bar{\boldsymbol{\Phi}}^H$ remains unchanged in $\boldsymbol{w}(n)$ but this component is orthogonal to $\boldsymbol{x}(n)$ so that the error signal $e(n)$ is not affected by this component. From (9), we note that

$$\boldsymbol{w}_{\text{opt}} = \bar{\boldsymbol{\Phi}}^H \boldsymbol{h}, \quad (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)\boldsymbol{h} = \boldsymbol{\psi}. \tag{31}$$

with $\boldsymbol{h} = [h_1, \ldots, h_K]^T$. Defining the new variable

$$z_i(n) = e^{-j\omega_i n}(h_i(n) - h_i) \quad (i = 1, \ldots, K) \tag{32}$$

(30) is written as

$$z_i(n+1) = e^{-j\omega_i} z_i(n) - \mu e^{-j\omega_i} \sum_{k=1}^{K} \sum_{l=1}^{K} A_i A_k^* e^{j(\omega_l-\omega_k)n} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l z_l(n). \tag{33}$$

The derivation is given in Appendix III. Defining the vectors $\boldsymbol{z}(n) = [z_1(n), \ldots, z_K(n)]^T$ and $\boldsymbol{b}_i(n) = e^{j\omega_i n}\boldsymbol{\phi}_i (i = 1, \ldots, K)$, (33) can be expressed as

$$\boldsymbol{z}(n+1) = \boldsymbol{\Lambda}_\omega^* \left[\boldsymbol{I} - \mu \boldsymbol{\pi}\boldsymbol{\pi}^H \boldsymbol{B}(n)\right] \boldsymbol{z}(n) \tag{34}$$

with

$$\boldsymbol{\Lambda}_\omega = \text{diag}(e^{j\omega_1}, \ldots, e^{j\omega_K}), \quad \boldsymbol{\pi} = [A_1 \ldots A_K]^T \tag{35}$$

$$\boldsymbol{B}(n) = \begin{bmatrix} \boldsymbol{b}_1^H(n) \\ \vdots \\ \boldsymbol{b}_K^H(n) \end{bmatrix} [\boldsymbol{b}_1(n), \ldots, \boldsymbol{b}_K(n)]. \tag{36}$$

It is interesting to note that for the case treated in [4], where complex sinusoids with known frequencies and unit amplitudes are used for input signals, the relevant linear system is time-invariant and is described by the same form as (34) with $\boldsymbol{\phi}, \boldsymbol{B}(n)$ replaced by $[1\ 1 \cdots 1]^T$ and $\boldsymbol{I}$, respectively.

Then, we note that

$$\boldsymbol{B}(n+1) = \boldsymbol{\Lambda}_\omega^* \boldsymbol{B}(n) \boldsymbol{\Lambda}_\omega. \tag{37}$$

To show that $\boldsymbol{z}(n)$ converges to $\mathbf{0}$ under a certain condition, we use the following Lyapunov function

$$V(n) = \boldsymbol{z}^H(n) \bar{\boldsymbol{B}}(n) \boldsymbol{z}(n). \tag{38}$$

From (36), we see that $\bar{\boldsymbol{B}}(n)$ is positive definite, so that $V(n) \geq 0$ and $V(n) = 0$ means $\boldsymbol{z}(n) = \mathbf{0}$. From (34) and (37), it can be shown that

$$V(n+1) - V(n) = -\mu \left(2 - \mu \boldsymbol{\pi}^H \boldsymbol{B}(n) \boldsymbol{\pi}\right) \left|\boldsymbol{r}^H(n)\boldsymbol{z}(n)\right|^2 \tag{39}$$

where

$$\boldsymbol{r}(n) = \boldsymbol{B}^H(n)\boldsymbol{\pi}. \tag{40}$$

The derivation is given in Appendix IV. Thus, under the condition

$$2 - \mu \boldsymbol{\pi}^H \boldsymbol{B}(n) \boldsymbol{\pi} > 0 \ \text{ for all } \ n > 0 \tag{41}$$

$V(n)$ decreases for $\boldsymbol{r}^H(n)\boldsymbol{z}(n) \neq 0$. From (2) and (36), (41) is written as

$$0 < \mu < \frac{2}{\sup_n \|\boldsymbol{x}(n)\|^2}. \tag{42}$$

If it happens that $V(n) = V(n+1) = \cdots = V(n+L)$, that is

$$\boldsymbol{r}^H(n)\boldsymbol{z}(n) = 0, \boldsymbol{r}^H(n+1)\boldsymbol{z}(n+1) = 0, \ldots, \boldsymbol{r}^H(n+L-1)\boldsymbol{z}(n+L-1) = 0$$

then from (34) it follows that

$$\boldsymbol{z}(n+i) = \boldsymbol{\Lambda}_\omega^* \boldsymbol{z}(n+i-1) \quad (i = 1, \ldots, L-1) \tag{43}$$

so that $\boldsymbol{z}(n+i) = (\boldsymbol{\Lambda}_\omega^*)^{i-1}\boldsymbol{z}(n)$. Hence, from this and (40) we have

$$\begin{bmatrix} \boldsymbol{\pi}^H \boldsymbol{B}(n) \\ \boldsymbol{\pi}^H \bar{\boldsymbol{B}}(n+1) \boldsymbol{\Lambda}_\omega^* \\ \vdots \\ \boldsymbol{\pi}^H \boldsymbol{B}(n+L-1) (\boldsymbol{\Lambda}_\omega^*)^{L-1} \end{bmatrix} \boldsymbol{z}(n) = 0. \tag{44}$$

From (37), we note $\boldsymbol{B}(n) = (\boldsymbol{\Lambda}_\omega^*)^n \boldsymbol{\Phi}\boldsymbol{\Phi}^H \boldsymbol{\Lambda}_\omega^n$. Substituting this into (44), we readily see that

$$\begin{bmatrix} \boldsymbol{\pi}^H \\ \boldsymbol{\pi}^H \boldsymbol{\Lambda}_\omega^* \\ \vdots \\ \boldsymbol{\pi}^H (\boldsymbol{\Lambda}_\omega^*)^{L-1} \end{bmatrix} (\boldsymbol{\Lambda}_\omega^*)^n \boldsymbol{\Phi}\boldsymbol{\Phi}^H \boldsymbol{\Lambda}_\omega^n \boldsymbol{z}(n) = 0.$$

This is rewritten as

$$\bar{\boldsymbol{\Phi}}^H \text{diag}\,(A_1^*, A_2^*, \ldots, A_K^*) (\boldsymbol{\Lambda}_\omega^*)^n \boldsymbol{\Phi}\boldsymbol{\Phi}^H \boldsymbol{\Lambda}_\omega^n \boldsymbol{z}(n) = \mathbf{0}.$$

But this means that $\boldsymbol{z}(n) = \mathbf{0}$, so $\boldsymbol{z}(n) \neq \mathbf{0}$ implies at least one of $\boldsymbol{r}^H(n+i)\boldsymbol{z}(n+i)$ $(i = 0, \ldots, L-1)$ is nonzero. Hence, $V(n+L) < V(n)$ as long as $\boldsymbol{z}(n) \neq \mathbf{0}$. This shows that $\boldsymbol{z}(n)$ converges to $\mathbf{0}$, that is, $\boldsymbol{w}(n)$ converges to $\boldsymbol{w}_{\text{opt}}$ under the condition (42).

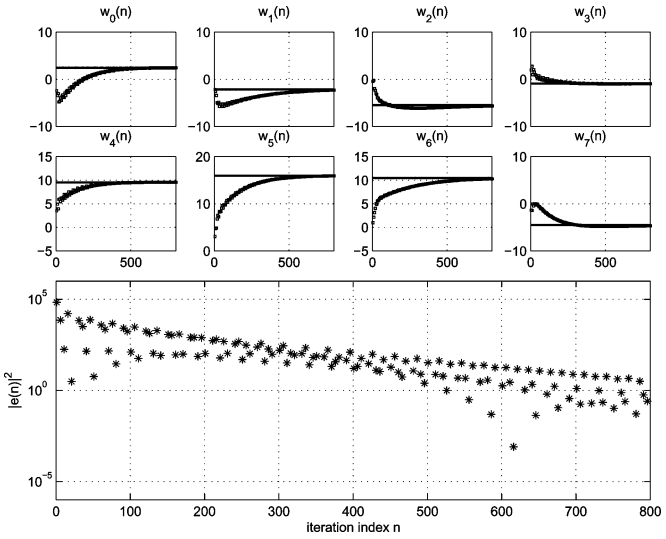


Fig. 2. Plots of tap weights and squared magnitude of error signals of LMS algorithm with $\mu = 0.0005$.

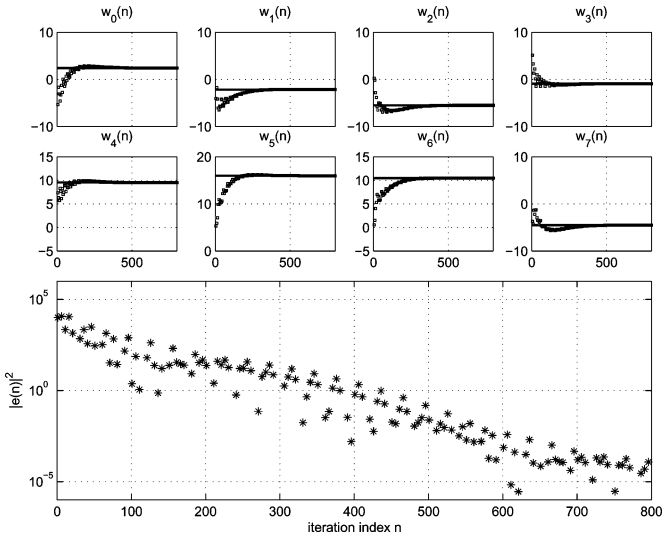


Fig. 3. Plots of tap weights and squared magnitude of error signals of LMS algorithm with $\mu = 0.0018$.

Remark: In [9, Sec. 6.1], it is shown that if $d(n) = \boldsymbol{w}_p^H \boldsymbol{x}(n)$, that is, $P(z)$ in Fig. 1 is of FIR structure, then $e(n) \to 0$ as $n \to \infty$ under (42) for general deterministic input signal. Moreover, if the PE condition is met, then $\boldsymbol{w}(n)$ converges to $\boldsymbol{w}_p$. In our special but important case of sinusoidal input signal with $L > K$, for any $\boldsymbol{w}_0$ satisfying (7), we have $d(n) = \boldsymbol{w}_0^H \boldsymbol{x}(n)$. Hence, we can apply the general result in [9]. However, this $\boldsymbol{w}_0$ is not unique and the PE condition is not satisfied, so that the limiting behavior of $\boldsymbol{w}(n)$ cannot be deduced from [9]. But by the previous development, we have shown that $\boldsymbol{w}(n)$ converges to $\boldsymbol{w}_{\text{opt}}$ in (9).

V. Simulation Results

To see the validity of the previous theoretical results, simulations are conducted using the LMS and RLS algorithm, respectively. In all the following simulations, the input signal $x(n)$ is real and is given by the sum of $4(K = 4)$ complex sinusoids with the amplitudes $A_1 = 1 + j$, $A_2 = 1 - j$, $A_3 = 2 + j$, $A_4 = 2 - j$, and the frequencies are $\omega_1 = \pi/3$, $\omega_2 = -\pi/3$, $\omega_3 = \pi/5$, $\omega_4 = -\pi/5$. The desired single $d(n)$ is formed by passing $x(n)$ through the stable transfer function $P(z) = 0.1 + 0.25z^{-1} - 2.9z^{-2} - 7.25z^{-3} + 10z^{-4} + 25z^{-5}$ and all the initial weights are set to zero.

First, we use the LMS algorithm to adjust the weights in the adaptive process. The FIR adaptive filter is assumed to have eight tap coefficients $w_i(n)$ $(i = 0, 1, \ldots, 7)$, that is, $L = 8$. From (9), the optimal weight vector is

$$\boldsymbol{w}_{\text{opt}} = [2.4052,\ -2.1428,\ -5.514,\ -0.92722,\ 9.5072,\ 15.952,\ 10.452,\ -4.5341]^T \quad (45)$$

and the upper bound of μ in (42) is $2/\sup_n ||\boldsymbol{x}(n)||^2 = 0.0019$. Figs. 2–4 show the convergence performance of the LMS adaptive filter. The step sizes are chosen as 0.0005, 0.0018, and 0.0038, respectively.

It can be seen from Fig. 2 that if the selected step size is too small, too many iterations are necessary for convergence. On the other hand, if it exceeds the upper bound, according to our exact analysis, the LMS algorithm cannot give a convergence result. We can see in Fig. 4 that the trajectories in this case are oscillating. If μ is increased further, the trajectories diverge exponentially. This is an interesting phenomenon in the linear time-varying system (34), where the oscillating region for μ has an interval of nonzero length whereas for usual linear time invariant systems such a region has length zero.

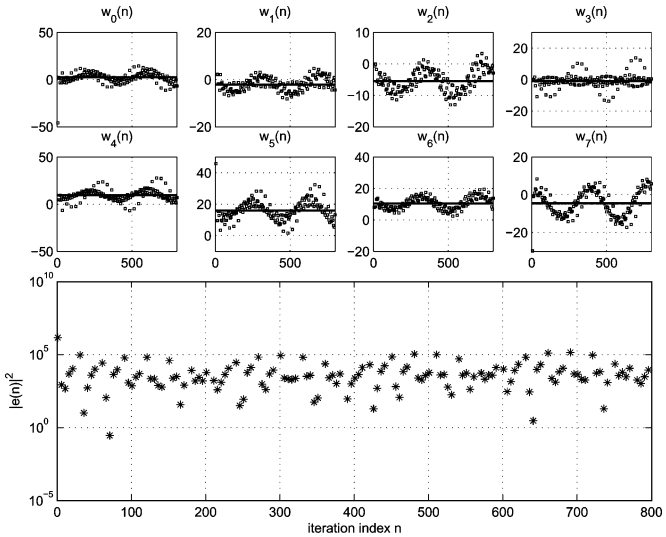


Fig. 4. Plots of tap weights and squared magnitude of error signals of LMS algorithm with $\mu = 0.0038$.

In the following, we simulate the performance of the LMS adaptive filter with different numbers of tap weights. Here, the number of coefficients in the transfer function $P(z)$ is 6. We assume the tap number of the FIR filter is $L = 6$, $L = 10$, or $L = 30$, respectively. For each choice of L, the corresponding step size limit given by (42) is 0.00289, 0.00203, and 0.00074, respectively. Here, we choose the step size to be close to the limits as $\mu = 0.0028$, $\mu = 0.002$, and $\mu = 0.0007$, respectively. It can be seen from Fig. 5 that the number of tap weights influences the convergence rate of the LMS adaptive filter considerably and the larger tap number produces the higher accuracy in this case. The explanation for this requires further studies.

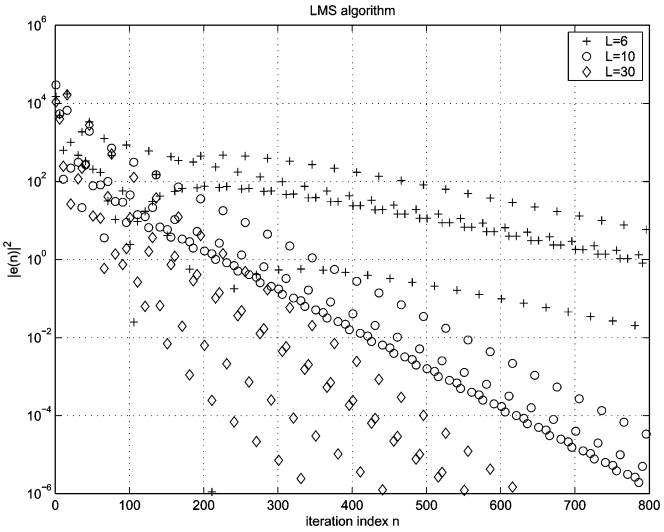


Fig. 5. Influence of the number of tap weights L on the performance of the LMS adaptive filter.

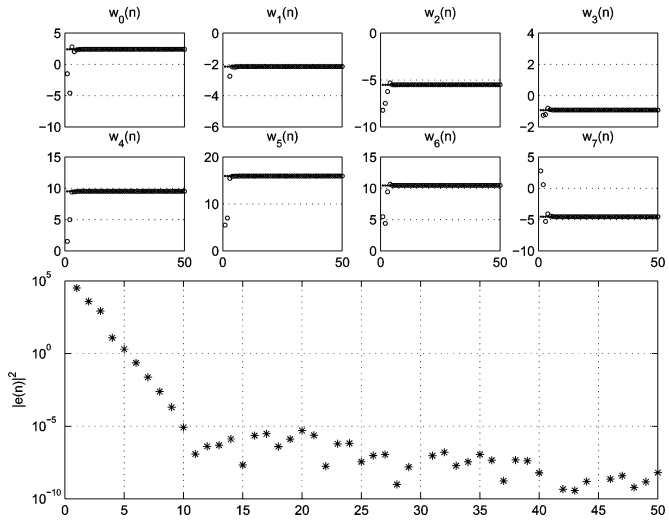


Fig. 6. Plots of tap weights and squared magnitude of error signals of RLS algorithm with the forgetting factor $\lambda = 0.9$.

Next, we use the RLS algorithm with forgetting factor $\lambda = 0.9$ to adjust the adaptive filter weights. As we can see from Fig. 6, the RLS algorithm converges to the same optimal weight shown in (45).

In Fig. 7, we treat the case where Gaussian white observation noise with variance 0.1 and mean zero is added to the desired signal $d(n)$. It is seen that the effect of the noise does not change the convergence property except that the error signal converges to the observation noise.

VI. Conclusion

In this paper, exact convergence analysis of the RLS and LMS algorithms has been presented for the case of sinusoidal noise canceling without the PE condition, where the reference signal to the adaptive filter is assumed to be a sum of K complex sinusoids with unknown amplitudes and unknown distinct frequencies. It is a future task to extend the present result to the case of the filtered-X LMS algorithm in active noise control.

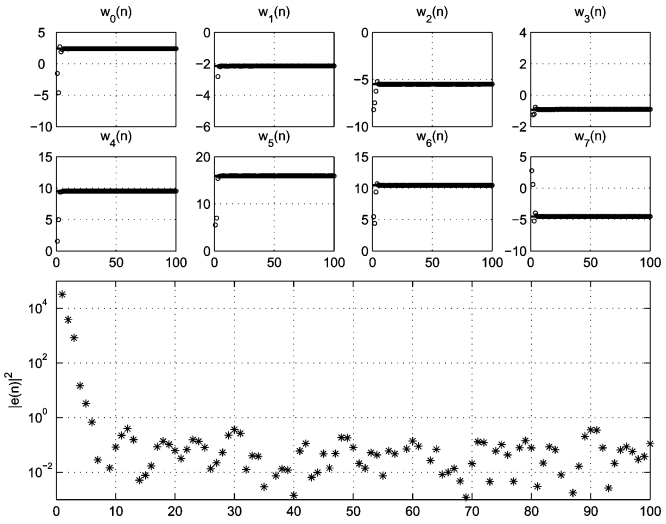


Fig. 7. Performance of RLS with observation noise and $\lambda = 1$.

Appendix I

In this appendix, we briefly state the PE condition.

The most general definition of the PE condition is defined in [9, Sec. 6.1] as follows.

A signal $\boldsymbol{x}(n)$ is strongly PE if there are positive constants α_1, α_2 and some window span N, such that

$$\alpha_1 \boldsymbol{I} \le \frac{1}{N} \sum_{n=k_0+1}^{k_0+N} \boldsymbol{x}(n)\boldsymbol{x}^H(n) \le \alpha_2 \boldsymbol{I}, \quad \text{for all } k_0 \ge 0 \quad (46)$$

where $\boldsymbol{x}(n) = (x(n)x(n-1), \ldots, x(n-L+1))^T$.

This means that the deterministic covariance matrix over a window span N is of rank L. But for the signal $\boldsymbol{x}(n)$ in (2), this rank is K for $L \ge K$ so that this matrix becomes rank deficient for $L > K$.

Appendix II

In this appendix, we derive (30).

Substituting (2), (5), and (29) into (28), we obtain

$$\sum_{i=1}^{K} h_i(n+1)\boldsymbol{\phi}_i = \sum_{i=1}^{K} h_i(n)\boldsymbol{\phi}_i + \mu \sum_{i=1}^{K} A_i \boldsymbol{\phi}_i e^{j\omega_i n} \times \left(\sum_{k=1}^{K} A_k^* P_k^* e^{-j\omega_k n} - \sum_{k=1}^{K} A_k^* \boldsymbol{\phi}_k^H e^{-j\omega_k n} \sum_{l=1}^{K} h_l(n)\boldsymbol{\phi}_l \right). \quad (47)$$

Since $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_k$ are linearly independent, we have

$$\begin{aligned} h_i(n+1) - h_i(n) &= \mu A_i e^{j\omega_i n} \times \left(\sum_{k=1}^{K} A_k^* P_k^* e^{-j\omega_k n} - \sum_{k=1}^{K} A_k^* \boldsymbol{\phi}_k^H e^{-j\omega_k n} \sum_{l=1}^{K} h_l(n)\boldsymbol{\phi}_l \right) \\ &= \mu \sum_{k=1}^{K} A_i A_k^* P_k^* e^{j(\omega_i - \omega_k)n} - \mu \sum_{k=1}^{K} A_i A_k^* e^{j(\omega_i - \omega_k)n} \times \sum_{l=1}^{k} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l(n) \end{aligned} \quad (48)$$

which yields (30).

APPENDIX III

In this appendix, we derive (33).

From (8) and (31), it is easy to see that

$$P_k^* = \sum_{l=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l \quad (k = 1, \ldots, K). \tag{49}$$

Performing the same operation on both sides of (30), we get

$$\begin{aligned} & e^{-j\omega_i(n+1)} \left(h_i(n+1) - h_i\right) \\ &= e^{-j\omega_i(n+1)} \left(h_i(n) - h_i\right) - \mu e^{-j\omega_i(n+1)} \\ &\quad \times \sum_{k=1}^{K} A_i A_k^* e^{j(\omega_i - \omega_k)n} \sum_{l=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l(n) \\ &\quad + \mu e^{-j\omega_i(n+1)} \sum_{k=1}^{K} A_i A_k^* P_k^* e^{j(\omega_i - \omega_k)n}. \end{aligned} \tag{50}$$

Substituting the definition of $z_i(n)$ given by (32) and applying (49) gives

$$\begin{aligned} & z_i(n+1) \\ &= e^{-j\omega_i} z_i(n) - \mu e^{-j\omega_i} \sum_{k=1}^{K} A_i A_k^* e^{-j\omega_k n} \sum_{l=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l(n) \\ &\quad + \mu e^{-j\omega_i} \sum_{k=1}^{K} A_i A_k^* e^{-j\omega_k n} \sum_{l=1}^{K} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l h_l \\ &= e^{-j\omega_i} z_i(n) - \mu e^{-j\omega_i} \\ &\quad \times \sum_{k=1}^{K} \sum_{l=1}^{K} A_i A_k^* e^{-j\omega_k n} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l \left(h_l(n) - h_l\right) \\ &= e^{-j\omega_i} z_i(n) - \mu e^{-j\omega_i} \\ &\quad \times \sum_{k=1}^{K} \sum_{l=1}^{K} A_i A_k^* e^{j(\omega_l - \omega_k)n} \boldsymbol{\phi}_k^H \boldsymbol{\phi}_l z_l(n). \end{aligned} \tag{51}$$

APPENDIX IV

In this appendix, we derive (39).

By (34), (37), and (38), it follows that

$$\begin{aligned} & V(n+1) - V(n) \\ &= \boldsymbol{z}^H(n+1)\boldsymbol{B}(n+1)\boldsymbol{z}(n+1) - \boldsymbol{z}^H(n)\boldsymbol{B}(n)\boldsymbol{z}(n) \\ &= \boldsymbol{z}^H(n) \left(I - \mu \boldsymbol{B}^H(n)\boldsymbol{\pi}\boldsymbol{\pi}^H\right) \boldsymbol{\Lambda}_\omega \boldsymbol{\Lambda}_\omega^* \boldsymbol{B}(n) \boldsymbol{\Lambda}_\omega \boldsymbol{\Lambda}_\omega^* \\ &\quad \times \left(I - \mu \boldsymbol{\pi}\boldsymbol{\pi}^H \boldsymbol{B}(n)\right) \boldsymbol{z}(n) - \boldsymbol{z}^H(n)\boldsymbol{B}(n)\boldsymbol{z}(n) \\ &= \boldsymbol{z}^H(n) \left(I - \mu \boldsymbol{B}^H(n)\boldsymbol{\pi}\boldsymbol{\pi}^H\right) \boldsymbol{B}(n) \\ &\quad \times \left(I - \mu \boldsymbol{\pi}\boldsymbol{\pi}^H \boldsymbol{B}(n)\right) \boldsymbol{z}(n) - \boldsymbol{z}^H(n)\boldsymbol{B}(n)\boldsymbol{z}(n) \\ &= -\mu \left(2 - \mu \boldsymbol{\pi}^H \boldsymbol{B}(n) \boldsymbol{\pi}\right) \left|\boldsymbol{r}^H(n)\boldsymbol{z}(n)\right|^2. \end{aligned} \tag{52}$$

The third equality follows since $\boldsymbol{\Lambda}_\omega \boldsymbol{\Lambda}_\omega^* = I$. The forth equality holds by $\bar{\boldsymbol{B}}^H(n) = \bar{\boldsymbol{B}}(n)$. The last equality follows from the definition of $\boldsymbol{r}(n)$ in (40).

REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[2] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, no. 6, pp. 484–491, Dec. 1977.

[3] M. Chakraborty and H. Sakai, "Convergence analysis of complex LMS algorithm with tonal reference signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 2, pp. 286–292, Mar. 2005.

[4] H. Sakai and Y. Hinamoto, "An exact analysis of the LMS algorithm with tonal reference signals in the presence of frequency mismatch," *Signal Process.*, vol. 85, pp. 1255–1262, 2005.

[5] L. Ljung, *System Identification*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[6] A. Albert, *Regression and the Moore-Penrose Pseudoinverse*. New York: Academic, 1972.

[7] B. Widrow and S. D. Stearns, *Adapt. Signal Process.*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[8] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. New York: Wiley, 1998.

[9] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

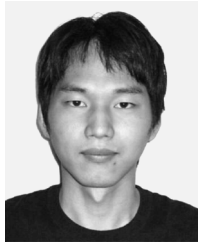
Hideaki Sakai (M'78–SM'02–F'07) received the B.E. and Dr. Eng. degrees in applied mathematics and physics from Kyoto University, Kyoto, Japan, in 1972 and 1981, respectively.

From 1987 to 1988, he spent six months at Stanford University, Stanford, CA, as a Visiting Scholar. He is currently a Professor at the Department of Systems Science, Graduate School of Informatics, Kyoto University. His research interests include the areas of statistical and adaptive signal processing.

Dr. Sakai served as an Associate Editor of IEICE *Transactions Fundamentals of Electronics, Communications, and Computer Sciences* from 1996 to 2000 and the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2001. He has been a member of the editorial board of EURASIP *Journal of Applied Signal Processing* from 2001 to 2005. He has been the chair of Japan Chapter of IEEE Signal Processing Society from 2005 to 2006.

Jun-Mei Yang (S'07) was born in Shandong, China, in 1979. She received the B.S. degree in applied mathematics from Shandong University, Jinan, China, in 2002, and the M.S. degree in cybernetics from the Chinese Academy of Sciences (CAS), Beijing, China, in 2005. She is currently pursuing the Ph.D. degree in mathematical system theory from the Graduate School of Informatics, Kyoto University, Kyoto, Japan.

Her research interests include theory and applications of independent component analysis, adaptive noise cancellation, and acoustic echo cancellation.



Tetsuya Oka was born in Osaka, Japan, in 1983. He received the B.E. degree in the informatics and mathematical science from Kyoto University, Kyoto, Japan, in 2005. He is currently pursuing the M.E. degree in adaptive signal processing at the Graduate School of Informatics, Kyoto University, Kyoto, Japan.