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 sinusoids 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 a 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 $\omega_k$ ($k = 1, \ldots, K$). For real signals, frequencies $\omega_k$ 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

$$
\mathbf{x}(n) = [x(n) \ x(n-1) \ \ldots \ x(n-L+1)]^T
$$

$$
\mathbf{w}(n) = [w_0(n) \ w_1(n) \ \ldots \ w_{L-1}(n)]^T
$$

(1)
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

$$x(n) = \sum_{k=1}^{K} A_k e^{j\omega_k n}, \quad x(n) = \sum_{k=1}^{K} A_k \phi_k e^{j\omega_k n}$$

(2) with

$$\phi_k = \begin{bmatrix} 1 & e^{-j\omega_k} & \cdots & e^{-j(L-1)\omega_k} \end{bmatrix}^T.$$  

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

$$y(n) = w^H(n)x(n) = \sum_{k=1}^{K} A_k w_k^H(n) \phi_k e^{j\omega_k n}$$

(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

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

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

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

$$\sum_{k=1}^{K} A_k w_k^H \phi_k e^{j\omega_k n} = \sum_{k=1}^{K} A_k P_k e^{j\omega_k n}$$

(6) so that $\mathbf{w}$ must satisfy

$$\mathbf{w} = \Phi^H \psi$$

(7) where

$$\Phi^H = [\phi_1 \phi_2 \cdots \phi_K]^H, \quad \psi = [P_1 P_2 \cdots P_K]^H$$

(8) and $\phi_k$ ($k = 1, \ldots, K$) are linearly independent for $L \geq K$. If $L < K$, (7) is an over determined system. In general, $\psi$ does not lie in the range space of $\Phi$, 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, $P_k = w_k^H \phi_k$ with the impulse response $w_k = (w_{p_1} \cdots w_{p_{L-1}})^T$. For this case (7) has a unique solution $w_k$. If $L = K$, obviously (7) has a unique solution. If $L > K$, (7) is an under-determined system whose solution is not unique. But $\Phi$ is of full row rank, so using the (Moore–Penrose) pseudo inverse $\Phi^+$ of $\Phi$, one solution is

$$w_{\Phi^+} = \Phi^+ \psi = \Phi^H(\Phi\Phi^H)^{-1}\psi$$

(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 $w_{\Phi^+}$ 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

$$Q(n)w(n) = q(n)$$

(10) where the deterministic covariance matrix $Q(n)$ and cross covariance vector $q(n)$ are defined by

$$Q(n) = \sum_{i=1}^{n} \lambda^{n-i} x(i)x^H(i)$$

(11) $$q(n) = \sum_{i=1}^{n} \lambda^{n-i} x(i)w^T(i)$$

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

$$Q(n) = \sum_{k=1}^{K} \sum_{l=1}^{K} A_k A_l^T \phi_k \phi_l^H D_n(\omega_k - \omega_l)$$

(12) $$q(n) = \sum_{k=1}^{K} \sum_{l=1}^{K} A_k A_l^T \phi_k D_n(\omega_k - \omega_l)$$

(13) where

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

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

$$Q(n) = (A\Phi^H)D(n)A\Phi, \quad q(n) = (A\Phi^H)D(n)A\psi$$

(15) where

$$A \triangleq \text{diag}(A_1^T, A_2^T, \ldots, A_K^T).$$

(16)
From (14) the following quadratic form:

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

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

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

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

$$D(n) = G^H(n)G(n)$$

(17)

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

$$Q(n) = (R(n)\Phi)^H R(n)\Phi, \quad q(n) = (R(n)\Phi)^H R(n)\psi$$

(18)

where \( R(n) = G(n)A \) is nonsingular. From (18), we note that \( Q(n) \) 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

$$u(n) = u(n-1) + \hat{Q}^{-1}(n) x(n)e^*(n)$$

$$e(n) = d(n) - u(n-1)^H x(n)$$

(19)

with

$$\hat{Q}(n) = \lambda \hat{Q}(n-1) + x(n)x^H(n)$$

(20)

and the initialization

$$u(0) = 0, \quad \hat{Q}(0) = \xi I \quad (\xi > 0)$$

(21)

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

$$\hat{Q}(n) = \lambda^n \xi I + Q(n).$$

(22)

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

$$u(n) = \hat{Q}^{-1}(n) q(n).$$

(23)

Let the eigenspace decomposition of \( Q(n) \) be

$$Q(n) = U \Lambda U^H$$

(24)

where for notational simplicity, we omit the dependence on \( n \) in the RHS of (24) and \( U = [u_1, \ldots, u_K, u_{K+1}, \ldots, u_L] \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K, 0, \ldots, 0) \) with \( u_i \) eigenvectors and \( \lambda_i \), the corresponding eigenvalues, respectively. From (12) and (13), it is obvious that \( u_{K+1}, \ldots, u_L \) corresponding to zero eigenvalues are orthogonal to \( \phi_1, \phi_2, \ldots, \phi_K \) and, hence, \( q(n) \). So, from (23) and (24), we have

$$u(n) = U (\lambda^n \xi I + \Lambda)^{-1} \text{diag}(1, \ldots, 1, 0, \ldots, 0) U^H q(n)$$

$$= U \text{diag} \left( \frac{1}{\lambda_1 + \lambda^n \xi}, \ldots, \frac{1}{\lambda_K + \lambda^n \xi}, 0, \ldots, 0 \right) U^H q(n).$$

(25)

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

$$Q^+(n) q(n) = ((R(n)\Phi)^H R(n)\Phi) + (R(n)\Phi)^H R(n)\psi$$

$$= (R(n)\Phi)^+ R(n)\psi$$

$$= (R(n)\Phi)^H (R(n)\Phi)^H R(n)\psi$$

$$= u_{\text{opt}}.$$  

(26)

That is, \( u(n) \) converges to \( u_{\text{opt}} \) in (9).

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

$$1/n q(n) = 1/n \sum_{i=1}^{L} x(i) d^*(i) + 1/n \sum_{i=1}^{L} x(i) \nu^*(i)$$

(27)

\( q(n)/n \) is still in the space spanned by \( \phi_1, \phi_2, \ldots, \phi_K \) and the covariance matrix of the second term in (27) is given by \( \sigma^2 \sum_{i=1}^{n} x(i)^2 n^2 \) for white noise \( \nu(n) \) with zero mean and variance \( \sigma^2 \). Since \( |x(n)| \approx \sqrt{(x(n)^2 x(n))} \) is bounded by some constant, this covariance matrix converges to zero and the effect of the observation noise fades out. But, of course, taking \( \lambda \) 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

$$u(n+1) = u(n) + \mu x(n) e^*(n)$$

$$e(n) = d(n) - u(n)^H x(n)$$

(28)

where \( \mu \) 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 $\mathbf{w}(0)$ is taken to $\mathbf{0}$, from (28) $\mathbf{w}(n)$ is written as

$$\mathbf{w}(n) = \sum_{i=1}^{K} h_i(n) \phi_i^o.$$  (29)

This may be regarded as a kind of discrete Fourier expansion though $\phi_i^o$'s are not orthogonal. From (2), (5), and (29), we have

$$h_i(n+1) = h_i(n) - \mu \sum_{k=1}^{K} A_i A_k^* e^{j\omega kn} \sum_{l=1}^{K} \phi_l^H \phi_l h_l(n) + \mu \sum_{k=1}^{K} A_i A_k^* P_k e^{j\omega kn} \left( i = 1, \ldots, K \right).$$  (30)

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

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

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

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

(30) is written as

$$z_i(n+1) = e^{-j\omega_i} z_i(n) - \mu e^{-j\omega_i} \sum_{k=1}^{K} A_i A_k^* e^{j(\omega_i-\omega_k) n} \phi_k^H \phi_i z_i(n).$$  (33)

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

$$\mathbf{z}(n+1) = \Lambda^o \left[ I - \mu \pi \Phi^H \mathbf{B}(n) \right] \mathbf{z}(n)$$  (34)

with

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

$$\mathbf{B}(n) = \begin{bmatrix} b_1^H(n) \\ \vdots \\ b_K^H(n) \end{bmatrix} [b_1(n), \ldots, b_K(n)].$$  (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 $\phi, \mathbf{B}(n)$ replaced by $[1 \ 1 \ \cdots \ 1]^T$ and $I$, respectively.

Then, we note that

$$\mathbf{B}(n+1) = \Lambda^o \mathbf{B}(n) \Lambda^o.$$  (37)

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

$$V(n) = \mathbf{z}^H(n) \mathbf{B}(n) \mathbf{z}(n).$$  (38)

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

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

where

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

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

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

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

$$0 < \mu < \frac{2}{\sup_{n} \left| \mathbf{r}(n) \right|^2}. $$  (42)

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

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

then from (34) it follows that

$$\mathbf{z}(n+i) = \Lambda^o \mathbf{z}(n+i-1) \quad (i = 1, \ldots, L - 1)$$  (43)

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

$$\begin{bmatrix} \pi H \mathbf{B}(n) \\ \pi H \mathbf{B}(n+1) \Lambda^o \\ \vdots \\ \pi H \mathbf{B}(n+L-1) (\Lambda^o)^{L-1} \end{bmatrix} \mathbf{z}(n) = \mathbf{0}.$$  (44)

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

$$\begin{bmatrix} \pi H \\ \pi H \Lambda^o \\ \vdots \\ \pi H (\Lambda^o)^{L-1} \end{bmatrix} (\Lambda^o)^n \Phi \Phi^H \Lambda^o \mathbf{z}(n) = \mathbf{0}.$$  (45)

This is rewritten as

$$\Phi^H \text{diag} (A_1^*, A_2^*, \ldots, A_K^*) (\Lambda^o)^n \Phi \Phi^H \Lambda^o \mathbf{z}(n) = \mathbf{0}.$$  (46)

But this means that $\mathbf{z}(n) = \mathbf{0}$, so $\mathbf{z}(n) \neq \mathbf{0}$ implies at least one of $\mathbf{r}^H(n+i) \mathbf{z}(n+i) (i = 0, \ldots, L - 1)$ is nonzero. Hence, $V(n+L) < V(n)$ as long as $\mathbf{z}(n) \neq \mathbf{0}$. This shows that $\mathbf{z}(n)$ converges to $\mathbf{0}$, that is, $\mathbf{w}(n)$ converges to $\mathbf{w}_{\text{opt}}$ under the condition (42).
Remark: In [9, Sec. 6.1], it is shown that if \( d(n) = w_p^T x(n) \), 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 \( \mathbf{w}(n) \) converges to \( \mathbf{w}_p \). In our special but important case of sinusoidal input signal with \( L > K \), for any \( \mathbf{w}_0 \) satisfying (7), we have \( d(n) = w_p^T x(n) \). Hence, we can apply the general result in [9]. However, this \( \mathbf{w}_0 \) is not unique and the PE condition is not satisfied, so that the limiting behavior of \( \mathbf{w}(n) \) cannot be deduced from [9]. But by the previous development, we have shown that \( \mathbf{w}(n) \) converges to \( \mathbf{w}_{opt} \) 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

\[
\mathbf{w}_{opt} = [2.4052, -2.1428, -5.514, -0.92722, 9.5072, 15.952, 10.452, -4.5341]^T \tag{45}
\]

and the upper bound of \( \mu \) in (42) is \( 2/\sup_{n} \|x(n)\|^2 = 0.0019 \). 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 \( \mu \) is increased further, the trajectories diverge exponentially. This is an interesting phenomenon in the linear time-varying system (34), where the oscillating region for \( \mu \) has an interval of nonzero length whereas for usual linear time-invariant systems such a region has length zero.

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 \( \mu = 0.0028, \mu = 0.002, \text{ and } \mu = 0.0007 \), 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.
It is seen that the effect of the noise does not change for
is of rank so that this matrix becomes rank de-
in (2), this such that
(46)

\[ \alpha_1 I \leq \frac{1}{N} \sum_{r=k_0+1}^{k_0+N} x(n)x^H(n) \leq \alpha_2 I, \quad \text{for all } k_0 \geq 0 \]

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

This means that the deterministic covariance matrix over a window span \( N \) is of rank \( L \). But for the signal \( x(n) \) in (2), this rank is \( K \) for \( L \geq K \) 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

\[
\sum_{k=1}^{K} h_k(n+1)\phi_k = \sum_{k=1}^{K} h_k(n)\phi_k + \mu \sum_{k=1}^{K} A_k\phi_k e^{j\omega_k n} \\
\times \left( \sum_{k=1}^{K} A_k^*P^e_k e^{-j\omega_k n} - \sum_{k=1}^{K} A_k^*\phi_k H e^{-j\omega_k n} \sum_{l=1}^{K} h_l(n)\phi_l \right). \tag{47}
\]

Since \( \phi_1, \ldots, \phi_k \) are linearly independent, we have

\[
h_k(n+1) - h_k(n) = \mu A_k e^{j\omega_k n} \\
\times \left( \sum_{k=1}^{K} A_k^*P^e_k e^{-j\omega_k n} - \sum_{k=1}^{K} A_k^*\phi_k H e^{-j\omega_k n} \sum_{l=1}^{K} h_l(n)\phi_l \right) \\
= \mu \sum_{k=1}^{K} A_k A_k^*P^e_k e^{j(\omega_k - \omega_l)n} - \mu \sum_{k=1}^{K} A_k A_k^* e^{j(\omega_k - \omega_l)n} \\
\times \sum_{l=1}^{K} \phi_l^H h_l(n) \tag{48}
\]

which yields (30).
APPENDIX III

In this appendix, we derive (33). From (8) and (31), it is easy to see that

\[ P_k = \sum_{l=1}^{K} \phi_k^H \phi_l h_l \quad (k = 1, \ldots, K). \]  

(49)

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

\[ e^{-j\omega_k (n+1)} (h_{l}(n+1) - h_{k}) = e^{-j\omega_k (n+1)} (h_{l}(n) - h_{k}) - \mu e^{-j\omega_k (n+1)} \]
\[ \times \sum_{l=1}^{K} A_l^* A_k e^{j(\omega_k - \omega_k)n} \sum_{l=1}^{K} \phi_k^H \phi_l h_l(n) \]
\[ + \mu e^{-j\omega_k (n+1)} \sum_{l=1}^{K} A_l^* A_k^w e^{j(\omega_k - \omega_k)n} \phi_k^H \phi_l z_l(n). \]  

(50)

Substituting the definition of \( z_l(n) \) given by (32) and applying (49) gives

\[ z_l(n+1) \]
\[ = e^{-j\omega_l} z_l(n) - \mu e^{-j\omega_l} \sum_{k=1}^{K} A_k^* A_k^w e^{-j\omega_l n} \sum_{l=1}^{K} \phi_k^H \phi_l h_l(n) \]
\[ + \mu e^{-j\omega_0} \sum_{k=1}^{K} A_k^* A_k^w e^{-j\omega_l n} \phi_k^H \phi_l (h_l(n) - h_l) \]
\[ \times \sum_{l=1}^{K} A_l^* A_k e^{j(\omega_k - \omega_l)n} \phi_k^H \phi_l z_l(n). \]  

(51)

APPENDIX IV

In this appendix, we derive (39).

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

\[ V(n+1) = V(n) \]
\[ = z^H(n+1) B(n+1) z(n+1) - z^H(n) B(n) z(n) \]
\[ = z^H(n) \left( I - \mu B^H(n) \pi^H \right) A_n A_n^* B(n) A_n A_n^* \]
\[ \times \left( I - \mu \pi B^H(n) \right) z(n) - z^H(n) B(n) z(n) \]
\[ = z^H(n) \left( I - \mu B^H(n) \pi^H \right) B(n) \]
\[ \times \left( I - \mu \pi B^H(n) \right) z(n) - z^H(n) B(n) z(n) \]
\[ = -\mu \left( 2 - \mu \pi B^H(n) \pi \right) |r^H(n) z(n)|^2. \]  

(52)

The third equality follows since \( A_n A_n^* = I \). The forth equality holds by \( B^H(n) = B(n) \). The last equality follows from the definition of \( r(n) \) in (40).