PAPER

Statistical-Mechanics Approach to Theoretical Analysis of the FXLMS Algorithm*

Seiji MIYOSHI[†a)] ***and*** **Yoshinobu KAJIKAWA**[††], ***Senior Members***

SUMMARY We analyze the behaviors of the FXLMS algorithm using a statistical-mechanical method. The cross-correlation between a primary path and an adaptive filter and the autocorrelation of the adaptive filter are treated as macroscopic variables. We obtain simultaneous differential equations that describe the dynamical behaviors of the macroscopic variables under the condition that the tapped-delay line is sufficiently long. The obtained equations are deterministic and closed-form. We analytically solve the equations to obtain the correlations and finally compute the mean-square error. The obtained theory can quantitatively predict the behaviors of computer simulations including the cases of both not only white but also nonwhite reference signals. The theory also gives the upper limit of the step size in the FXLMS algorithm.

key words: *FXLMS algorithm, adaptive filter, active noise control, statistical-mechanical informatics, long-filter assumption*

1. Introduction

The adaptive filter is one of the major techniques in digital signal processing. Its purpose is to identify an unknown system through which a reference signal propagates. Active noise control (ANC), which has been practically realized owing to the progress of digital signal processing technology [1]–[4], is an application of the adaptive filter. ANC is divided into two types, feedforward and feedback ANC [3], [4]. In this paper, we focus on the feedforward ANC. In the ANC system, it should be noted that a secondary path also exists, which is the propagating path from the canceling loudspeaker to the error microphone.

The least-mean-square (LMS) algorithm is the most commonly used algorithm in adaptive filters. The LMS algorithm was proposed more than half a century ago [5]–[7]. When the LMS algorithm is applied to ANC, the secondary path should be estimated beforehand and a signal that has passed through the estimated secondary path is used. We call this procedure the Filtered-X LMS (FXLMS) algorithm [8]. This algorithm is a generalization of the LMS algorithm because the FXLMS algorithm is equivalent to the LMS algorithm when the secondary path impulse response is a single unit pulse with no delay in time domain.

To theoretically analyze the behaviors of the LMS algorithm, various methods have been proposed. One of the most commonly used methods is the independence assumption [9]–[11], and the FXLMS algorithm has also been analyzed on the basis of this assumption [12]–[15]. In a finite-duration impulse response (FIR) filter, which is commonly used as the adaptive filter [6], [7], the elements of the tap input vector are shifted along the tapped-delay line. Theoretical treatment of the effect of such a tap input vector is intractable. Professor Haykin [6] expressed this difficulty as "Indeed, despite the extensive effort that has been expended in the literature to analyze the LMS filter, we still do not have a direct mathematical analysis of its stability and steady-state performance, and quite probably we never will." To evade this difficulty, with the independence assumption, successive tap input vectors of the tapped-delay line are assumed to be independently generated at each time step. However, the actual elements of the tap input vector are merely shifted to the next position. Hence, each tap input vector is related to the previous one and the vectors are thus not independent. Owing to this fact, analyses based on the independence assumption involve essential and potential problems [6], [7]. Note that this assumption is incorrect even if the reference signal is independently generated, that is, the reference signal is white.

Various previous studies have been based on assumptions other than the independence assumption. In [16]–[19], another form of independence was assumed. That is, the correlation between the tap input vectors is assumed to be greater than the correlation between the coefficient vector of the adaptive filter and the tap input vectors. However, analytical results naively based on such assumptions cannot precisely predict experimental results, particularly when there is little or no background noise. In [15], [20] and [21], it was assumed that the step size is small. In [22], [23] and [24], it was assumed that the reference signal is sinusoidal. Thus, a general theory for the FXLMS algorithm has not been given in the literature even though these algorithms are widely used.

Meanwhile, statistical mechanics, which is a branch of physics, was constructed with the aim of understanding the macroscopic properties of gases and magnets from the properties of microscopic elements such as molecules and atoms. In statistical mechanics, numerous powerful ana-

Manuscript received May 1, 2018.
Manuscript revised July 31, 2018.

[†]The author is with the Department of Electrical, Electronic and Information Engineering, Faculty of Engineering Science, Kansai University, Suita-shi, 564-8680 Japan.

[††]The author is with the Department of Electrical and Electronic Engineering, Faculty of Engineering Science, Kansai University, Suita-shi, 564-8680 Japan.

*The contents of this paper were partially presented at 2012 IEEE Statistical Signal Processing Workshop (SSP2012), 2013 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2013), and 2015 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2015).

a) E-mail: miyoshi@kansai-u.ac.jp

DOI: 10.1587/transfun.E101.A.2419

lytical and numerical methods have been developed. The field in which these methods are used to solve problems in information technology or information science is called statistical-mechanical informatics [25], which is producing significant results in many fields, such as associative memory models [26], [27], error-correcting codes [28], wireless communications [29], image processing [30], statistical learning [31], and so forth. In statistical-mechanical informatics, universal properties, which are independent of the realizations of individual problems, are macroscopically discussed by assuming the large-system limit. In regard to statistical learning, numerous models have been analyzed, in which learning machines are modeled using perceptrons [25], [31].

Comparing an FIR filter with a linear perceptron that has a linear output function, it can be observed that they resemble each other in that their outputs are the sums of products of many inputs and coefficients. Focusing on this point, in this paper we theoretically analyze the behaviors of the FXLMS algorithm by applying a statistical-mechanical method that has been used to analyze on-line learning [25], [31]–[33]. Cross-correlations between the elements of a primary path and those of an adaptive filter and autocorrelations of the elements of the adaptive filter are treated as the macroscopic variables. We obtain simultaneous differential equations that describe the dynamical behaviors of the macroscopic variables under the condition that the tapped-delay line is sufficiently long. The obtained equations are deterministic and closed-form. We analytically solve the equations to obtain the correlations and finally compute the mean-square error (MSE). That is, we tackle the intractable problem by using the assumptions that the correlation between the coefficient vector of the adaptive filter and the tap input vectors is small and that the impulse responses of the primary path and the adaptive filter are sufficiently long. Note that this long-impulse-response assumption or long-filter assumption used in our analysis is reasonable, considering actual acoustic systems. The obtained theory quantitatively agrees with the results of computer simulations as shown later.

The authors previously reported an approximate version [34] of the theory based on a statistical-mechanical method. However, the approximate version cannot predict simulation results when the step size is large since it neglects the correlations, which are generated by the effect of the secondary path, between past tap input vectors and the coefficient vector of the adaptive filter. In this paper, we derive an exact theory taking the above correlations into consideration and describe details of the analysis to show the effectiveness of the statistical-mechanical method in the field of signal processing. In contrast to the approximate version [34], the obtained theory is valid even when the step size is large since it computes the correlations between past tap input vectors and the coefficient vector of the adaptive filter. This breakthrough also makes it possible to obtain the upper bound of the step size as described in this paper.

The analysis considering the above correlations was partially reported in [35]. However, for analytical convenience in [35] as well as in [34], the primary path was a series connection of an unknown system and the secondary path in the model. This assumption was also a severe limitation. Therefore, in this paper we obtain a theory in which the primary path does not include the secondary path. Furthermore, our previous theories [34], [35] treated only a white reference signal. This assumption was also a severe limitation since a reference signal is not necessarily white. Therefore, in this paper, we generalize the previous theories to the case where the reference signal is not necessarily white by introducing the correlation function of the reference signal. In addition, we show that the theory can predict the behaviors in the case of an actual primary path by absorbing the autocorrelations of the primary path. These generalizations were partially reported in [36] and [37], respectively. However, in this paper, we present the entire theory, which has been further generalized regarding the tap lengths of the primary path and adaptive filter, and we discuss the validity and limitations of the theory in detail. Furthermore, the steady-state analysis is discussed in this paper.

This paper is organized as follows. After describing the analytical model considered in this paper in Sect. 2, the theory of the FXLMS algorithm is derived in Sect. 3 using a statistical-mechanical method. In Sect. 4, the validity of the derived theory is verified by comparison with simulation results. In addition, the learning curves under several conditions, the steady-state analysis, and the effect of the estimation error of the secondary path are discussed. Conclusions are presented in Sect. 5.

2. Analytical Model

Figure 1 shows a block diagram of the feedforward ANC system considered in this paper. The primary path P is represented by an N_p-tap FIR filter. Its coefficient vector is

$$\boldsymbol{p} = [p_1, p_2, \ldots, p_{N_p}]^T, \tag{1}$$

where ^{T} denotes the transpose of a vector. Each coefficient p_i is assumed to be generated from the stochastic process expressed as

$$\langle p_i \rangle = 0, \qquad \left\langle p_i p_{i-j} \right\rangle = \kappa_j, \tag{2}$$

and is time-invariant. Here, $\langle \cdot \rangle$ denotes expectation. If $\kappa_j = 0$ when $j \neq 0$ in (2), the elements of the impulse response of the primary path are uncorrelated with each other. In this paper, we consider the autocorrelated primary path by introducing the covariance function κ_j.

The adaptive filter H is an N_h-tap FIR filter. Its coefficient vector is

$$\boldsymbol{h}(n) = [h_1(n), h_2(n), \ldots, h_{N_h}(n)]^T, \tag{3}$$

where n denotes the time step. The initial value $h_i(0)$ of each coefficient is assumed to be generated from the stochastic process expressed as

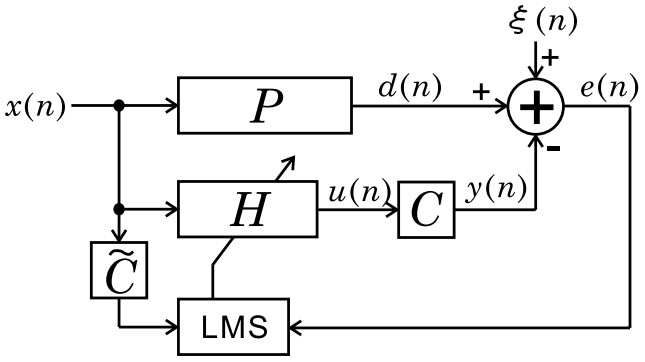

Fig. 1 Block diagram of the feedforward ANC system.

$$\langle h_i(0)\rangle = 0, \qquad \left\langle h_i(0)h_j(0)\right\rangle = \delta_{i,j}. \tag{4}$$

Here, δ denotes the Kronecker delta, defined as $\delta_{i,j} = 1$ ($i = j$), or 0 (otherwise). Note that the case where $h_i(0)$ are all zero can be analyzed in the framework of the present paper by a slight modification of the initial value of the differential equations as described in Sect. 3.

The reference signal $x(n)$ is assumed to be drawn from a distribution with

$$\langle x(n)\rangle = 0, \qquad \langle x(n)x(n-k)\rangle = \frac{r_k}{N_h}. \tag{5}$$

The reference signal $x(n)$ is shifted through the tapped-delay line. Therefore, the tap input vectors of the primary path P and adaptive filter H are

$$\boldsymbol{x}_p(n) = [x(n), x(n-1), \ldots, x(n-N_p+1)]^T \tag{6}$$

and

$$\boldsymbol{x}_h(n) = [x(n), x(n-1), \ldots, x(n-N_h+1)]^T, \tag{7}$$

respectively. In the above model, $\|\boldsymbol{p}\|$ and $\|\boldsymbol{h}(0)\|$ are $O(\sqrt{N_p})^\dagger$ and $O(\sqrt{N_h})$, respectively. On the other hand, $\|\boldsymbol{x}_p(n)\|$ and $\|\boldsymbol{x}_h(n)\|$ are $O(1)$. Here, $\|\cdot\|$ denotes the ℓ_2-norm. The covariance of the reference signal $x(n)$ is normalized by N_h in (5). Assuming such a covariance and normalizing the time step n by N_h, we can compare the learning curves of different N_h as described later. Note that generality is not lost by this normalization because we can determine the value of r_k such that r_k/N_h equals the actual covariance. For example, if the actual variance of the reference signal is 0.1 and the actual number N_h of taps of the adaptive filter H is 200, we only have to set $r_0 = 20$. Considering the limit $N_p, N_h \to \infty$ as described later, we can analytically obtain dynamical behaviors that are independent of N_p and N_h. It should also be noted that only the mean and covariance of the distribution are defined in (5). No specific distributions, for example, the Gaussian distribution, are assumed. The same statements are true for both the coefficients of the primary path P and the initial values of the coefficients of the adaptive filter H. If $r_k = 0$ when $k \neq 0$ in (5), the reference signal is white. In this paper, we analyze the behavior of the FXLMS algorithm for a nonwhite reference signal by introducing the covariance function r_k.

Since both P and H are FIR filters, their outputs are convolutions of their own coefficients and a sequence of reference signals. That is, the outputs $d(n)$ of the primary path P and $u(n)$ of the adaptive filter H are

$$d(n) = \boldsymbol{p}^T\boldsymbol{x}_p(n), \qquad u(n) = \boldsymbol{h}(n)^T\boldsymbol{x}_h(n), \tag{8}$$

respectively.

The secondary path C is modeled by a K-tap FIR filter. Its coefficient vector is

$$\boldsymbol{c} = [c_1, c_2, \ldots, c_K]^T \tag{9}$$

and is time-invariant. Since C is also an FIR filter, its output is a convolution of its coefficients and its input sequence. Therefore, the output $y(n)$ of the secondary path C is

$$y(n) = \sum_{k=1}^{K} c_k u(n-k+1). \tag{10}$$

The error signal $e(n)$ is generated by adding an independent background noise $\xi(n)$ to the difference between $d(n)$ and $y(n)$. That is,

$$e(n) = d(n) - y(n) + \xi(n). \tag{11}$$

Here, the mean and variance of $\xi(n)$ are zero and σ_ξ^2, respectively.

The LMS algorithm is used to update the adaptive filter. Here, the coefficient vector $\boldsymbol{c}$ of the secondary path C is unknown in general. Therefore, the estimated secondary path $\tilde{C}$, which has been estimated in advance by a certain method, is used to update the adaptive filter. This procedure is called the FXLMS algorithm. When the estimated secondary path $\tilde{C}$ is a K-tap FIR filter and its coefficient vector is

$$\tilde{\boldsymbol{c}} = [\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_K]^T, \tag{12}$$

the update procedure obtained by the FXLMS algorithm is

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu e(n) \sum_{k=1}^{K} \tilde{c}_k \boldsymbol{x}_h(n-k+1), \tag{13}$$

where μ is the step size.

3. Theory

In this section, we theoretically analyze the behaviors of the FXLMS algorithm using a statistical-mechanical method. From (10) and (11), the MSE is expressed as follows [34]–[37].

$$\left\langle e^2(n)\right\rangle = \left\langle d^2(n)\right\rangle$$

†We use the big O notation throughout this paper. We express $f(N) = O(g(N))$ if and only if there exists a positive real number L and a real number N_0 such that $|f(N)| \le L|g(N)|$ for all $N > N_0$.

$$+\sum_{k=1}^{K}\sum_{k'=1}^{K} c_k c_{k'} \langle u(n-k+1)u(n-k'+1)\rangle$$
$$-2\sum_{k=1}^{K} c_k \langle d(n)u(n-k+1)\rangle + \sigma_\xi^2 \tag{14}$$

Equation (14) includes many means of the products of u and u and the products of d and u, including cases where their time steps are different. To calculate these products, we introduce the N_h-dimensional vectors

$$\boldsymbol{k}_j(n) = [k_{j,1}(n), k_{j,2}(n), \ldots, k_{j,N_h}(n)]^T. \tag{15}$$

Here, $j = -M, \ldots, M$ and the elements of the vectors are

$$k_{j,i}(n) = h_{\mathrm{mod}(i+j-1,N_h)+1}(n), \tag{16}$$

where $\mathrm{mod}(i+j-1, N_h)$ denotes the remainder when $i+j-1$ is divided by N_h. Therefore, (15) and (16) indicate that shifts of up to M are considered. This means that the correlations for up to M shifts are considered as described later. Note that $\boldsymbol{k}_0(n) = \boldsymbol{h}(n)$. Here, the fact that $\boldsymbol{k}_j(n)$ is the *shifted* vector of $\boldsymbol{h}(n)$ is important. Although (16) expresses $\boldsymbol{k}_j(n)$ as the circularly shifted vector of $\boldsymbol{h}(n)$, circular shifting is not essential. In the long-filter assumption described below, generality is not lost even if we assume a linearly shifted vector instead of a circularly shifted vector because the effect of the ends of $\boldsymbol{h}(n)$ and $\boldsymbol{k}_j(n)$ can be ignored macroscopically since these vectors are infinitely long in the long-filter assumption.

In the following, long filters, that is, $N_p, N_h \to \infty$, are assumed. This condition corresponds to *the thermodynamic limit* in statistical mechanics[25]. With this assumption, we can deterministically describe the macroscopic behaviors of the system, as described later, although the reference signal is stochastically generated. The long-filter assumption is more acceptable than the independence assumption [9]–[15] used in the literature since the numbers of taps of the unknown system and the adaptive filter are usually large in actual acoustic systems.

If the shift number j is $O(1)$, we obtain

$$\boldsymbol{h}(n)^T \boldsymbol{x}_h(n) = \boldsymbol{k}_j(n)^T \boldsymbol{x}_h(n-j). \tag{17}$$

Equation (17) is justified in Appendix A. Equation (17) is based on the fact that the shift of the tap input vector is canceled by the shift of the elements of the adaptive filter. Here, the effect of the ends of the adaptive filter can be ignored since both $\boldsymbol{h}(n)$ and $\boldsymbol{k}_j(n)$ are N_h-dimensional, i.e., infinitely long, vectors. Equation (17) implies that the shift j in the time direction can be replaced by the subscript of vector $\boldsymbol{k}$. In addition, we consider that

$$a = \frac{N_p}{N_h} \tag{18}$$

is kept constant and introduce two macroscopic variables $R_j(n)$ and $Q_j(n)$ defined by

$$R_j(n) = \frac{1}{\bar{a}N_h}\sum_{i=1}^{\bar{a}N_h} p_i k_{j,i}(n), \tag{19}$$

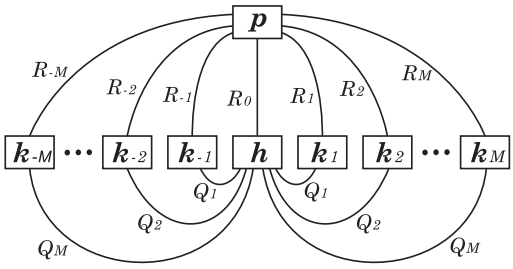

Fig. 2 Relationships among $\boldsymbol{p}, \boldsymbol{h}, \boldsymbol{k}$, and correlations R and Q.

$$Q_j(n) = \frac{1}{N_h}\sum_{i=1}^{N_h} h_i(n) k_{j,i}(n). \tag{20}$$

Here, $\bar{a}$ is the smaller value of a and 1, that is,

$$\bar{a} = \min(a, 1). \tag{21}$$

Considering (16), it can be seen that $R_j(n)$ and $Q_j(n)$ are equal to the means of $p_i h_{i+j}(n)$ and $h_i(n)h_{i+j}(n)$, respectively. That is, $R_j(n)$ and $Q_j(n)$ are the cross-correlation between $\boldsymbol{p}$ and $\boldsymbol{h}(n)$ and the autocorrelation of $\boldsymbol{h}(n)$, respectively. Figure 2 shows the relationships among the vectors and the correlations. Note that $Q_j(n)$ is symmetric, that is, $Q_j(n) = Q_{-j}(n)$, from the definitions of $Q_j(n)$ and $\boldsymbol{k}_j(n)$. $Q_0(n)$ is the mean of $h_i^2(n)$ and is omitted in Fig. 2.

Then, we obtain

$$\langle d(n-j)u(n)\rangle = \bar{a}\sum_{i=-M}^{M} R_i(n) r_{i-j}, \tag{22}$$

$$\langle u(n-j)u(n)\rangle = \sum_{i=-M}^{M} Q_i(n) r_{i-j}, \tag{23}$$

$$\langle d(n-j)d(n)\rangle = a\sum_{i=-L}^{L} \kappa_i r_{i-j}. \tag{24}$$

Equations (22) and (24) are derived in detail in Appendix B and Appendix C, respectively. As described in Appendix B, we derived (22) based on the assumption that the correlation between $\boldsymbol{h}(n)$ and $\boldsymbol{x}_h(n)$ is small [16]–[18]. As described in Sect. 1, theoretical treatment of the LMS algorithm is intractable [6]. Therefore, we tackle this difficult problem by using the above small-correlation assumption and the long-filter assumption. Equation (23) is obtained in the same manner as (22). Equations (22) and (23) indicate that the correlations for up to only M shifts are considered. Therefore, the $2M+1$ vectors $\{\boldsymbol{k}_j\}$, $j = -M, \ldots, M$ are considered and it is assumed that $R_j = Q_j = 0$ when $|j| > M$. The larger the value of M, the larger the computational cost and the more precisely the theory can predict the simulation results as shown later. In (24), L is the range in which two elements of the primary path have some correlation. From (22)–(24), we can express the MSE (14) in terms of $R_j(n)$ and $Q_j(n)$ as

$$\left\langle e^2(n)\right\rangle = \sum_{k=1}^{K} c_k \sum_{i=-M}^{M}\left(\sum_{k'=1}^{K} c_{k'} Q_i(n) r_{i-k+k'}\right.$$

$$- 2\bar{a}R_i(n)r_{i+k-1}\Big) + a\sum_{i=-L}^{L}\kappa_i r_i + \sigma_\xi^2. \tag{25}$$

This formula shows that the MSE is a function of the macroscopic variables $R_j(n)$ and $Q_j(n)$. Therefore, we derive differential equations that describe the dynamical behaviors of these variables in the following.

We first derive a differential equation for $R_j(n)$. When the coefficient vector $\boldsymbol{h}(n)$ of the adaptive filter is updated, the j-shifted vector $\boldsymbol{k}_j(n)$ is also changed. This change can be described as

$$\boldsymbol{k}_j(n+1) = \boldsymbol{k}_j(n) + \mu e(n)\sum_{k=1}^{K}\tilde{c}_k\boldsymbol{x}_h(n-k+1-j). \tag{26}$$

Note that the time step of the tap input vector $\boldsymbol{x}_h$ in (26) is shifted by j compared with that in (13). We define the N_h-dimensional vector $\bar{\boldsymbol{p}}$ as

$$\bar{\boldsymbol{p}} = [p_1, p_2, \ldots, p_{\bar{a}N_h}, \underbrace{0, \ldots, 0}_{(1-\bar{a})N_h}]^T. \tag{27}$$

Multiplying both sides of (26) on the left by $\bar{\boldsymbol{p}}^T$ and using (20), we obtain

$$\bar{a}N_hR_j(n+1) = \bar{a}N_hR_j(n) + \mu e(n)\sum_{k=1}^{K}\tilde{c}_k\bar{d}(n-k+1-j). \tag{28}$$

Here, $\bar{d}(n) = \bar{\boldsymbol{p}}^T\boldsymbol{x}_h(n)$. Then, we obtain

$$\left\langle \bar{d}(n-j)u(n)\right\rangle = \langle d(n-j)u(n)\rangle, \tag{29}$$

$$\left\langle \bar{d}(n-j)d(n)\right\rangle = \frac{\bar{a}}{a}\langle d(n-j)d(n)\rangle. \tag{30}$$

In (28), the left-hand side and the first term on the right-hand side are $O(N_h)$ and the other term is $O(1)$. This means that the coefficient vector $\boldsymbol{h}(n)$ of the adaptive filter should be updated $O(N_h)$ times to change $R_j(n)$ by $O(1)$. Therefore, we introduce time $t = n/N_h$ and use it to represent the adaptive process. Setting the variance and covariance of the reference signal $x(n)$ to be $O(N_h^{-1})$ as in (5) and considering the limit $N_h \to \infty$, we can deterministically describe the system's behavior using a small number of macroscopic variables as follows. Then, t becomes a continuous variable since the limit $N_h \to \infty$ is considered. This is a standard method in the statistical-mechanical analysis of on-line learning [31].

If the adaptive filter is updated $\lfloor N_h dt\rfloor$ times in an infinitely small time dt, we can obtain $\lfloor N_h dt\rfloor$ equations as follows:

$$\bar{a}N_hR_j(n+1) = \bar{a}N_hR_j(n) + \mu e(n)\sum_{k=1}^{K}\tilde{c}_k\bar{d}(n-k+1-j), \tag{31}$$

$$\bar{a}N_hR_j(n+2) = \bar{a}N_hR_j(n+1) + \mu e(n+1)\sum_{k=1}^{K}\tilde{c}_k\bar{d}(n-k+2-j), \tag{32}$$

$$\vdots \quad \vdots \quad \vdots$$

$$\bar{a}N_hR_j(n+\lfloor N_hdt\rfloor) = \bar{a}N_hR_j(n+\lfloor N_hdt\rfloor - 1) + \mu e(n+\lfloor N_hdt\rfloor - 1)\times\sum_{k=1}^{K}\tilde{c}_k\bar{d}(n-k+\lfloor N_hdt\rfloor - j), \tag{33}$$

where $\lfloor N_h dt\rfloor$ denotes the largest integer not greater than $N_h dt$. Summing all these equations, we obtain

$$\bar{a}N_h(R_j + dR_j) = \bar{a}N_hR_j + \lfloor N_hdt\rfloor\mu \times\left\langle e(m)\sum_{k=1}^{K}\tilde{c}_k\bar{d}(m-k+1-j)\right\rangle. \tag{34}$$

Here, from the law of large numbers, we have represented the effect of the probabilistic variables by their means since the updates are executed $\lfloor N_h dt\rfloor$ times, that is, many times, to change R_j by dR_j. This property is called *self-averaging* in statistical mechanics [25]. In (31)–(33), the difference in the time steps of e and $\bar{d}$ is $-k+1-j$. To express this, an auxiliary time step m has been introduced in (34). From (10), (11), (22), (24), (29), (30), and (34), we obtain a differential equation that describes the dynamical behavior of R_j in a deterministic form as follows:

$$\frac{dR_j}{dt} = \mu\sum_{k'=1}^{K}\tilde{c}_{k'}\Big(\sum_{i=-L}^{L}\kappa_i r_{i+k'+j-1} - \sum_{k=1}^{K}c_k\sum_{i=-M}^{M}R_i r_{i+k-k'-j}\Big). \tag{35}$$

Next, multiplying (13) by (26) and proceeding in the same manner as for the derivation of the above differential equation for R_j, we can derive a differential equation for Q_j, which is given by (36), where $\mathrm{sgn}(\cdot)$ and $\Theta(\cdot)$ are the sign function and step function defined by

$$\mathrm{sgn}(x) = 1\ (x \geq 0),\ \text{or}\ -1\ (\text{otherwise}), \tag{37}$$

$$\Theta(x) = 1\ (x \geq 0),\ \text{or}\ 0\ (\text{otherwise}), \tag{38}$$

respectively. In addition,

$$\alpha = \Theta(\gamma)\gamma - k'', \qquad \beta = \Theta(\epsilon)\epsilon - k'', \tag{39}$$

$$\gamma = j + 1 - k', \qquad \epsilon = -j + 1 - k'. \tag{40}$$

Here, we have to treat the correlations carefully and execute complicated calculations since the past tap input vectors $\boldsymbol{x}_h$ affect both the coefficient vector $\boldsymbol{h}$ of the adaptive filter and its shifted vector $\boldsymbol{k}_j$. The derivation of (36) is outlined in Appendix D.

Equations (35) and (36) are first-order ordinary differential equations with $3M+2$ variables since the correlations for up to M shifts are considered as described before. That is,

$$\frac{d}{dt}\boldsymbol{z} = \boldsymbol{G}\boldsymbol{z} + \boldsymbol{b}, \tag{41}$$

$$\frac{dQ_j}{dt} = \mu \sum_{k'=1}^{K} \tilde{c}_{k'} \Bigg\{ \sum_{i=-M}^{M} \Big[\bar{a} R_i \left(r_{i-\gamma} + r_{i-\epsilon} \right) - \sum_{k=1}^{K} c_k Q_i \left(r_{i-k'+k+j} + r_{i-k'+k-j} \right) \Big]$$
$$- \mu \Bigg[\mathrm{sgn}(\gamma) \sum_{k''=1}^{|\gamma|} \Bigg(\delta_{\alpha,0} \sigma_\xi^2 + a \sum_{i=-L}^{L} \kappa_i r_{i+\alpha}$$
$$- \sum_{k=1}^{K} c_k \sum_{i=-M}^{M} \left(\bar{a} R_i r_{i+k-1-\alpha} + \bar{a} R_i r_{i+k-1+\alpha} - \sum_{k'''=1}^{K} c_{k'''} Q_i r_{i-k+k'''-\alpha} \right) \Bigg) \sum_{i=1}^{K} \tilde{c}_i r_{k'-i-j+\alpha}$$
$$+ \mathrm{sgn}(\epsilon) \sum_{k''=1}^{|\epsilon|} \Bigg(\delta_{\beta,0} \sigma_\xi^2 + a \sum_{i=-L}^{L} \kappa_i r_{i+\beta}$$
$$- \sum_{k=1}^{K} c_k \sum_{i=-M}^{M} \left(\bar{a} R_i r_{i+k-1-\beta} + \bar{a} R_i r_{i+k-1+\beta} - \sum_{k'''=1}^{K} c_{k'''} Q_i r_{i-k+k'''-\beta} \right) \Bigg) \sum_{i=1}^{K} \tilde{c}_i r_{k'-i+j+\beta} \Bigg] \Bigg\}$$
$$+ \mu^2 \Bigg[\sum_{k=1}^{K} c_k \sum_{i=-M}^{M} \left(\sum_{k'=1}^{K} c_{k'} Q_i r_{i-k+k'} - 2 \bar{a} R_i r_{i+k-1} \right) + a \sum_{i=-L}^{L} \kappa_i r_i + \sigma_\xi^2 \Bigg] \sum_{k''=1}^{K} \sum_{k'''=1}^{K} \tilde{c}_{k''} \tilde{c}_{k'''} r_{k''-k'''-j}, \tag{36}$$

$$z = [Q_0, \ldots, Q_M, R_{-M}, \ldots, R_0, \ldots, R_M]^T. \tag{42}$$

Here, the matrix $\boldsymbol{G}$ and vector $\boldsymbol{b}$ are determined by (35) and (36). For example, the $(2M+2+j, 2M+2+i)$ element of the matrix $\boldsymbol{G}$ is equal to the coefficient of R_i on the right-hand side of (35), that is,

$$G_{ji} = -\mu \sum_{k'=1}^{K} \tilde{c}_{k'} \sum_{k=1}^{K} c_k r_{i+k-k'-j}, \tag{43}$$

where $-M \le i, j \le M$. Note that R_0 is the $(2M+2)$th element of z as shown in (42). As another example, the $(2M+2+j)$th element of the vector $\boldsymbol{b}$ is

$$b_j = \mu \sum_{k'=1}^{K} \tilde{c}_{k'} \sum_{i=-L}^{L} \kappa_i r_{i+k'+j-1} \tag{44}$$

as seen on the right-hand side of (35), where $-M \le j \le M$.

All initial values of $Q_j (j \neq 0)$ and R_j are equal to zero because $h_i(0)$ and p_i are independently generated. The initial value of Q_0 is unity since the initial value $h_i(0)$ of each coefficient is generated from the stochastic process given by (4). Therefore, z at $t = 0$ is

$$z(0) = [1, \underbrace{0, \ldots, 0}_{3M+1}]^T. \tag{45}$$

Using (45) as the initial condition, we can analytically solve (41) to obtain

$$z(t) = e^{\boldsymbol{G}t} \left(z(0) - \boldsymbol{G}^{-1} e^{-\boldsymbol{G}t} \boldsymbol{b} + \boldsymbol{G}^{-1} \boldsymbol{b} \right), \tag{46}$$

where $e^{\boldsymbol{D}}$ is the matrix exponential function defined by

$$e^{\boldsymbol{D}} = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{D}^k = \boldsymbol{I} + \frac{1}{1!} \boldsymbol{D} + \frac{1}{2!} \boldsymbol{D}^2 + \cdots. \tag{47}$$

If we consider the case where the initial values of the coefficients $h_i(0)$ are all zero, we can obtain the solution by modifying (45) to

$$z(0) = [\underbrace{0, \ldots, 0}_{3M+2}]^T \tag{48}$$

since the initial value of Q_0 is zero in this case.

The averaging method and the ordinary differential equation (ODE) method are well-known methods for analyzing the dynamical behaviors of discrete-time update systems [7], [21], [38], [39]. Here, we describe the difference between these methods and our proposed method. When the averaging method and ODE method are applied, continuous-time differential equations are derived from discrete-time equations based on the theorem that the dynamical behaviors of the variables correspond to the average behaviors in the small-step-size limit. Therefore, the small-step-size limit is essential in the averaging method and ODE method. In contrast, in our proposed method, the small-step-size condition is not explicitly used. In this case, continuous-time differential equations are derived through the reasonable processes described in this section. Here, we note the following. Although the variance and covariance of the reference signal are normalized by N_h, as shown in (5), this normalization is not equivalent to the small-step-size assumption. This is because the magnitude of the output of the adaptive filter is kept to the same order when N_h approaches infinity since not only the variance and covariance of the reference signal become small as shown in (5) but also the length of the adaptive filter becomes large. Therefore, the theory derived in this section is also expected to predict the experimental behaviors of a system in which neither the step size nor the covariance of the reference signal is infinitely small. To show the validity of the theory, in Sect. 4 we compare and discuss the results obtained using the theory and simulation results for a primary path and an adaptive filter with finite numbers of taps when neither the step size nor

the covariance of the reference signal is infinitely small.

4. Results and Discussion

In this section, we discuss the validity of the theory obtained in the previous section by comparison with simulation results. The conditions of the analysis and the simulation in the following subsections may not appear to be consistent. This is because each set of conditions is specifically selected depending on the purpose of each subsection. The main conditions in each subsection are summarized in Table 1.

4.1 Learning Curves and Ranges of Correlations

We first investigate the validity of the theory by comparison with simulation results regarding the dynamical behaviors of the MSE, that is, the learning curves. Figure 3 shows the learning curves obtained using the theory in the previous section, along with the corresponding simulation results. The conditions are shown in Table 1. The step size is $\mu = 0.1$. The parameters that determine the variance and covariance of the elements of the primary path are $\kappa_k = \delta_{k,0}$, that is, the elements of the primary path are uncorrelated. The parameters that determine the variance and covariance of the reference signal are $r_k = \delta_{k,0}$, that is, the reference signal is white. There is no background noise, that is, $\sigma_\xi^2 = 0$. In Fig. 3, the curves represent theoretical results and the polygonal lines represent simulation results. In the theoretical calculation, the results are obtained by substituting R_j and Q_j, which are obtained by solving (41), into (25) in the cases where the ranges of the correlations considered are $M = 0, 1, 2, 6$, and 10. Here, (41) is numerically solved by the Runge-Kutta method because the calculation of the analytical solution (46) involves numerical instability[††]. In the computer simulations, four given primary paths are used in the both cases of the numbers of taps of the primary path and the adaptive filter are $N_p = N_h = 80$ and 2000. For all the computer simulations, ensemble means for 200 trials are plotted for each given primary path. Each coefficient p_i of all the primary paths is independently generated from the Gaussian distribution with a mean of zero and variance of unity. Here, note that all of the simulation results using $N_p = N_h = 80$ and 2000 and the theoretical results derived using $N_p, N_h \to \infty$ can be compared along a single vertical axis because the variance of the reference is normalized by N_h, as shown in (5), and the variable $t = n/N_h$ is introduced instead of the number of updates n itself as described before. The steady-state MSE, as well as the learning curves, is independent of N_h since the variance of the reference signal is normalized by N_h in the model assumed in this paper. All initial values of the coefficients $h_i(0)$ are set to zero in the simulation, and the initial condition (48) is used in the theoretical calculation to compare the derived theory with the previous theory in [19]. The reason why the theory in [19] was chosen for comparison is that it is a relatively new theory that avoids the independence assumption and is used to analyze the dynamical behaviors of the MSE, similarly to the theory derived in this paper.

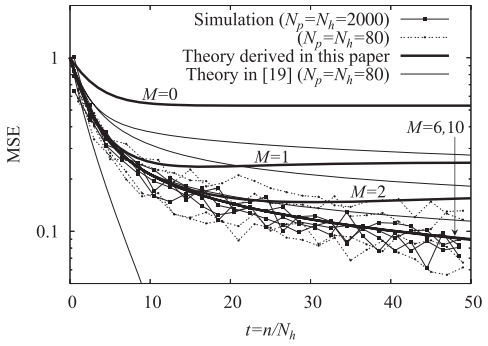

Fig. 3 Learning curves obtained theoretically and by simulation. The theoretical results for $M = 6$ and $M = 10$ almost overlap.

Figure 3 shows that the learning curves obtained by the computer simulations considerably vary in the case of $N_p = N_h = 80$, whereas the variation is small in the case of $N_p = N_h = 2000$. The theory derived in this paper predicts the simulation results well in the average sense. The larger the value of M, the more closely the theoretical results agree with the average simulation results. The theoretical results for $M = 6$ and $M = 10$ almost overlap and agree with the average simulation results reasonably well. The long-time dynamical behaviors can be predicted by using a large value of M.

The four learning curves obtained from the theory in [19], which assumes a small step size and statistical independence between the adaptive filter and the tap input vectors, are shown with thin lines in Fig. 3. In each of the four cases, exactly the same primary path as that used in the computer simulations for $N_p = N_h = 80$ has been used. Note that the theory in [19] does not provide an asymptotic analysis. Therefore, the results depend on the realized concrete values of the primary path. The theoretical results obtained by the theory in [19] excessively vary and are too sensitive to the given primary path. The theory cannot predict the corresponding simulation results in the model setting described in this paper. On the other hand, the theory derived in this paper predicts the average simulation results reasonably well.

The theory derived in this paper can also predict simulation results for large N_p and N_h, i.e., 2000, using a relatively small M, that is, using a small number of macroscopic variables, which is a noteworthy characteristic of the derived theory. It is the fundamental principle of statistical mechanics that the behavior of a system composed of a very large number of elements can be represented by a small number of macroscopic variables. Therefore, the expression 'large N_h', which is not quantified and appears imprecise, is not inappropriate. Note that Fig. 3 shows that the derived theory can predict average simulation results not only for large N_p

[††]The numerical instability regarding the analytical solution originates from the fact that the analytical solution includes the matrix exponential function defined by (47). Since (47) is expressed as an infinite series, some numerical computations are necessary and instability exists.

Table 1 Summary of the main conditions.

Sect.	Fig.	μ	σ_ξ^2	r_k	κ_k	K	$\boldsymbol{c}$	$\tilde{\boldsymbol{c}}$	Theory	Simulation	
									M	N_p, N_h	Ens.
4.1	3	0.1	0	$\delta_{k,0}$	$\delta_{k,0}$	2	$[1,1]^T$	$\leftarrow$	0–10	80 2000	200
4.2	4	0.1	0	$\delta_{k,0}$	$\delta_{k,0}$	25	(fn. 6)	$\leftarrow$	20	100	4000
4.3	5	0.1–0.4	0	$\delta_{k,0}$ $r_0=1, r_{\pm1}=0.5$	$\delta_{k,0}$	2	$[1,1]^T$	$\leftarrow$	20	500	1000
4.4	6	0.05–0.6	0.1	$\delta_{k,0}$	$\delta_{k,0}$	2	$[1,1]^T$	$\leftarrow$	50	1000 4000	1000
4.5	8	0.1	10^{-5}	$\delta_{k,0}$	(fn. ††††)	2	$[1,1]^T$	$\leftarrow$	10	256	1000
4.6	9	0.1	0–0.4	$\delta_{k,0}$	$\delta_{k,0}$	1 2 3	1 $[1,1]^T$ $[1,1,1]^T$	$\leftarrow$ $\leftarrow$ $\leftarrow$	20	200	200
4.7	10	0.1	0	$\delta_{k,0}$	$\delta_{k,0}$	2	$[1,1]^T$	$[1,\tilde{c}_2]^T$	100	200	50

μ: step size,
σ_ξ^2: variance of the background noise,
r_k: parameter that decides the autocorrelation of the reference signal,
κ_k: covariance of the elements of the primary path,
K: number of taps of the secondary path C,
$\boldsymbol{c}$: coefficients of the secondary path C,
$\tilde{\boldsymbol{c}}$: coefficients of the estimated secondary path $\tilde{C}$,
M: range in which the macroscopic variables R_j and Q_j are considered in the proposed theory,
N_p: number of taps of the primary path P in the simulation,
N_h: number of taps of the adaptive filter H in the simulation,
Ens.: number of trials for which the ensemble means are calculated in the simulation,
fn.: footnote,
δ: Kronecker delta.
$\tilde{c}_2$ is varied from -1 to $+3$ in Sect. 4.7.

and N_h, i.e., 2000, but also for relatively small N_p and N_h, i.e., 80.

There is some background (BG) noise in actual situations where ANC is applied. Previous studies [12] and [19] also treated the case where there is some BG noise. However, when there is BG noise, the accuracy of the theory is unclear since the MSE is buried in the BG noise. In contrast, much higher accuracy is needed when attempting to use the theory to predict the simulation results in the case of no BG noise. Therefore, to verify the theory, it is more rigorous and reasonable to show the agreement with simulation results in the case of no BG noise. For this reason, the results for a case with no BG noise are shown in Fig. 3. Of course, the proposed theory is also valid when there is BG noise, as shown later in Figs. 6, 8, and 9.

4.2 Long Impulse Response of Secondary Path

Figure 3 shows the results for $K = 2$, that is, the secondary path has a small number of taps. Next, we investigate the case of a large K to show the effectiveness of the obtained theory in a more realistic case. Figure 4 shows the result for $K = 25$. In both the theoretical calculation and the computer simulations, the coefficient vector $\boldsymbol{c}$ of the secondary path is the normalized version of the vector reported in [2][†††]. The initial coefficients $h_i(0)$ are independently generated from the Gaussian distribution with a mean of zero and variance of unity in the simulation, and the initial condition (45) is used in the theoretical calculation. Ensemble means in the simulation are taken by averaging over not only the reference signals but also the stochastically generated primary path. The other conditions are shown in Table 1. Figure 4 shows that the theory also predicts the average simulation results when the number of taps of the secondary path K is large.

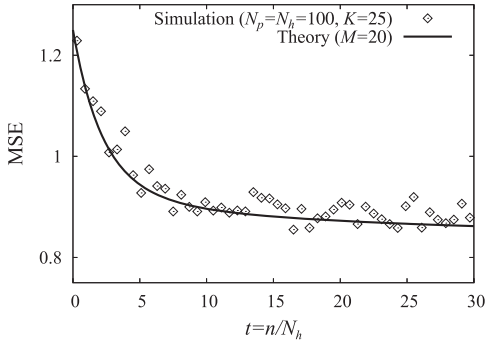

Fig. 4 Learning curves obtained theoretically and by simulation when $K = 25$.

4.3 Nonwhite Reference Signal

Figures 3 and 4 show the results for white reference signals. However, the reference signal is not necessarily white in the actual application of an ANC system. Therefore, we next investigate the case of a nonwhite reference signal

[†††] $\boldsymbol{c}$ = [−0.073844, 0.280609, −0.778907, −0.321869, 2.563927, −5.876712, 11.504314, −18.114344, 25.967108, −33.871861, 41.345085, −47.070049, 50.785450, −51.863720, 50.043129, −45.990845, 39.786583, −33.078777, 25.692730, −18.861567, 12.781915, −8.174841, 4.567914, −2.076808, 0.728670]^T

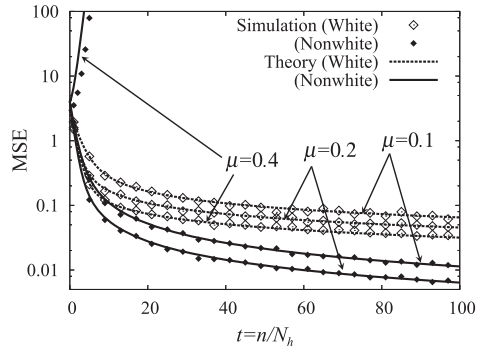

Fig. 5 Learning curves obtained theoretically and by simulation when the reference signal is white and nonwhite.

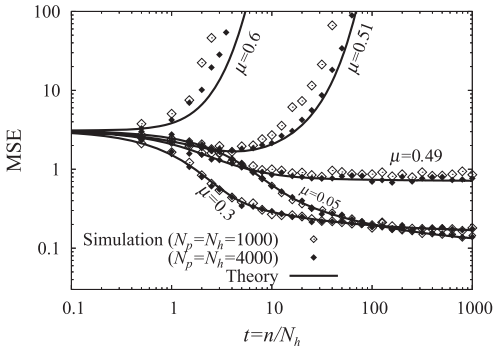

Fig. 6 Learning curves for various step sizes μ obtained theoretically and by simulation.

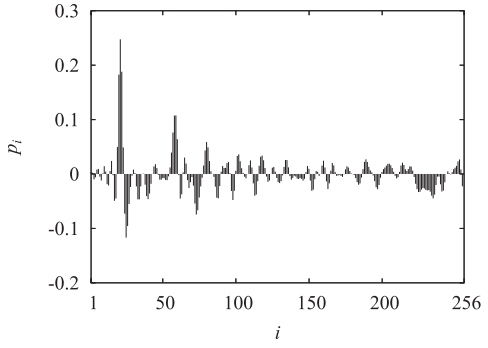

Fig. 7 Example of impulse response of an actual primary path.

to show the effectiveness of the obtained theory. Figure 5 shows the learning curves obtained theoretically along with the corresponding simulation results [36]. The correlation functions of the reference signal are $r_k = \delta_{k,0}$ (white) and $r_0 = 1, r_{\pm 1} = 0.5, r_{\pm k} = 0$ when $k \geq 2$ (nonwhite). The initial coefficients $h_i(0)$ are independently generated from the Gaussian distribution with a mean of zero and variance of unity in the simulation, and the initial condition (45) is used in the theoretical calculation. Ensemble means in the simulation are taken by averaging over not only the reference signals but also the stochastically generated primary path. The other conditions are shown in Table 1.

Figure 5 shows that the theoretical results agree with the simulation results including the difference between the behaviors for white and nonwhite reference signals. It is also shown that the upper bound of the step size μ for the white reference signal is larger than 0.4, whereas that for the nonwhite reference signal is smaller than 0.4 because the learning curve of $\mu = 0.4$ for the nonwhite reference signal diverges.

4.4 Effect of Step Size on Learning Curves

We next investigate the relationship between the step size μ and the learning curves. This is important from the viewpoint of the actual application of ANC systems.

Figure 6 shows the learning curves obtained theoretically along with the corresponding simulation results for various step sizes μ. The variance of the background noise is $\sigma_\xi^2 = 0.1$. The initial coefficients $h_i(0)$ are independently generated from the Gaussian distribution with a mean of zero and variance of unity in the simulation, and the initial condition (45) is used in the theoretical calculation. Ensemble means in the simulation are taken by averaging over not only the reference signals but also the stochastically generated primary path. The other conditions are shown in Table 1.

Figure 6 shows that the theoretical results agree with the simulation results reasonably well when μ is smaller than 0.5. However, there is significant disagreement when $\mu \geq 0.5$. The simulation results for $N_p = N_h = 4000$ more closely agree with the theoretical results than those for $N_p = N_h = 1000$. The reason for this disagreement is considered to be that the theory is derived using the long-filter assumption, that is, $N_p, N_h \to \infty$, whereas the simulations are executed using finite values of N_p and N_h. The dependence of the bias on the system size is an example of the phenomenon known as *the finite-size effect* in statistical mechanics. The condition of $\mu = 0.5$ corresponds to the phase transition point at which the MSE changes from being convergent to divergent. Figure 6 shows that this theory quantitatively predicts the simulation results for a finite number of taps when the step size is smaller than the transition point, which is usually the case for the FXLMS algorithm.

4.5 Actual Primary Path

In the derivation of the theory, the elements p_i are assumed to be stochastically generated as shown in (2). In addition, Fig. 3 shows the results for $\kappa_k = \delta_{k,0}$, that is, the case where the elements of the primary path P are uncorrelated. However, the actual elements of the impulse response of a primary path are not stochastic but deterministic. In addition, the elements are generally strongly correlated with each other, that is, the primary path has a strong autocorrelation. For example, Fig. 7 shows an example of the impulse response of a primary path obtained experimentally [37]††††. While the variance of the elements of the impulse response

††††The autocorrelations of this primary path are $\hat{\kappa}_j \equiv \frac{1}{N_P - j} \sum_{i=1+j}^{N_P} p_i p_{i-j}$ are (1.211, 0.963, 0.392, −0.134, −0.363,

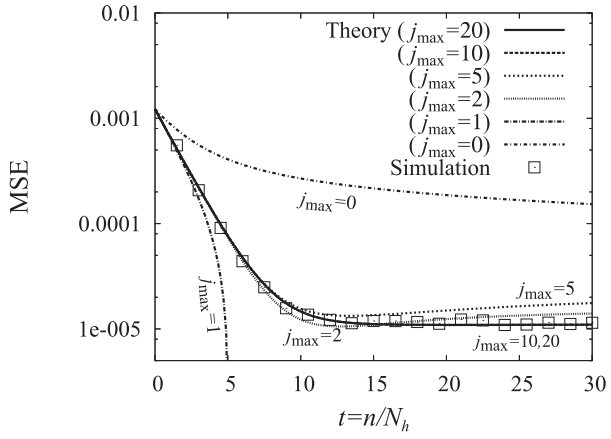

Fig. 8 Learning curves obtained theoretically and by simulation in the case of the actual primary path.

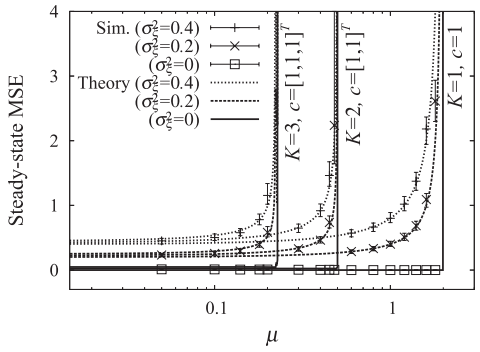

Fig. 9 Relationship between step size μ and steady-state MSE obtained theoretically and by simulation.

of this primary path is $\hat{\kappa}_0 \equiv \frac{1}{N_P}\sum_{i=1}^{N_P} p_i^2 = 1.211 \times 10^{-3}$, the correlation between the neighboring elements is $\hat{\kappa}_1 \equiv \frac{1}{N_P-1}\sum_{i=2}^{N_P} p_i p_{i-1} = 0.963 \times 10^{-3}$. This indicates that there are strong correlations between the elements of the impulse response of the actual primary path. Therefore, it is important to investigate whether the derived theory can predict the behaviors of computer simulations using the impulse response of the actual primary path. For this purpose, the theory can absorb the characteristics of the primary path by setting $\kappa_j = \hat{\kappa}_j$, where κ_j is defined by (2).

Figure 8 shows the learning curves obtained theoretically along with the corresponding simulation results in the case of the actual primary path [37].

In the theoretical calculations, the results of six cases are plotted, that is, $j_{\max} = 0, 1, 2, 5, 10$, and 20. Here, $j_{\max}$ is the maximum value of the subscript j of κ_j defined by (2). For example, $j_{\max} = 0$ means that $\kappa_j = 0, |j| > 0$, that is, the theory only considers the variance of the elements of the impulse response of the primary path. $j_{\max} = 2$ means that $\kappa_j = 0, |j| > 2$, that is, the theory considers $\kappa_j, j = -2, -1, 0, 1, 2$. In the computer simulations, $N_h = N_p = 256$ and ensemble means for 1000 trials are plotted. The impulse response of the primary path in the computer simulation is that shown in Fig. 7, that is, the actual impulse response measured experimentally. Note that the ensemble means are obtained by averaging over only reference signals. This averaging is different from that in the simulations in Sects. 4.1–4.4, where the ensemble means are obtained by averaging over both reference signals and channels. The other conditions are shown in Table 1.

Figure 8 shows that the theory for $j_{\max} = 0$, that is, the theory only using the variance, cannot predict the behaviors of the computer simulation. The theory for $j_{\max} = 1$ also cannot predict the behaviors. However, the larger the value of $j_{\max}$, the more closely the theoretical results agree with the simulation results. The theoretical results for $j_{\max} = 10$ and 20 almost overlap and agree with the simulation results reasonably well.

As described in this section, by using the autocorrelations of the primary path, the derived theory can predict the behaviors in the case of an actual primary path. This shows that even if the realized values of the actual elements, which are inherently deterministic, of the impulse response of the primary path are not known, the theory can predict the behaviors in the case of an actual primary path through the statistic $\hat{\kappa}$.

4.6 Steady-State MSE

We next investigate the steady-state MSE on the basis of the derived theory. Steady-state values of R and Q can be calculated by letting the left-hand side of (41) be a zero vector. That is, the steady-state value $\boldsymbol{z}_{st}$ of $\boldsymbol{z}$ is

$$\boldsymbol{z}_{st} = -\boldsymbol{G}^{-1}\boldsymbol{b}. \tag{49}$$

Substituting the obtained steady-state R and Q, that is, the elements of $\boldsymbol{z}_{st}$, into (25), we analytically obtain the steady-state MSE. We compare the theoretically obtained steady-state MSE and simulation results. The conditions regarding the number K of taps and the coefficient vector $\boldsymbol{c}$ of the secondary path C are $K = 1$ ($\boldsymbol{c} = 1$), $K = 2$ ($\boldsymbol{c} = [1, 1]^T$), and $K = 3$ ($\boldsymbol{c} = [1, 1, 1]^T$). The variances of the background noise are $\sigma_\xi^2 = 0, 0.2$, and 0.4. Figure 9 shows the results obtained theoretically and by simulation [35]. In the computer simulations, the medians and standard deviations of the 200 squared errors from $t = 9000$ to $t = 11000$ in 200 trials are plotted using error bars. These values of t in the simulations are sufficient for the MSE to reach a steady-state value. Ensemble means in the simulation are taken by averaging over not only the reference signals but also the stochastically generated primary path. The other conditions are shown in Table 1. The theoretical results agree with the simulation results reasonably well in Fig. 9.

4.7 Effect of Estimation Error of Secondary Path

Figures 3–9 show results in the case of $\tilde{\boldsymbol{c}} = \boldsymbol{c}$, i.e., the estimated secondary path has no error. However, the estimated

−0.312, −0.161, −4.41, 0.0156, 0.0406, 0.0284, −0.0243, −0.0757, −0.0680, −0.00586, 0.0290, −0.0362, −0.177, −0.287, −0.286, −0.184) × 10^{-3}.

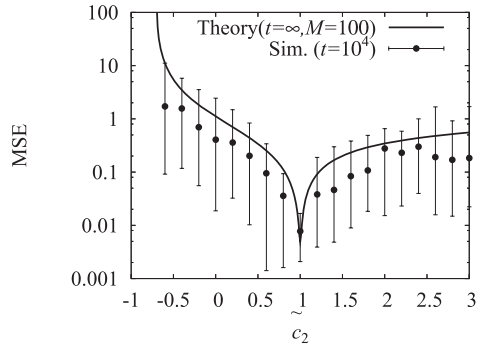

Fig. 10 Relationship between estimated secondary path and steady-state MSE.

secondary path inevitably has some error in actual active noise control. Therefore, it is important to investigate its effect. The obtained theory is also applicable to cases where the secondary path estimation has some error since the theory treats the secondary path and the estimated secondary path separately.

Figure 10 shows the relationship between the estimated secondary path and the steady-state MSE obtained theoretically along with the corresponding simulation results. The secondary path C is a two-tap FIR filter, that is, $K = 2$, and its coefficients are $c_1 = c_2 = 1$. The first element of the estimated secondary path is $\tilde{c}_1 = 1$ and the second element is varied from $\tilde{c}_2 = -1$ to $+3$. The step size is $\mu = 0.1$. In the computer simulations, the means and standard deviations of 50 squared errors in the case of $t = 10000$ are plotted using error bars. In the theoretical calculation, the steady-state MSE is obtained by the calculation described in Sect. 4.6. Ensemble means in the simulation are taken by averaging over not only the reference signals but also the stochastically generated primary path. The other conditions are shown in Table 1.

When $\tilde{c}_2 = 1$, which corresponds to no estimation error, the steady-state MSE is minimized. The larger the estimation error, the larger the steady-state MSE. The theoretical results qualitatively agree with the tendency of the simulation results in Fig. 10.

The so-called 90° condition [40]–[42] is well known for stability of the feedforward ANC. According to the 90° condition, the phase error between the secondary path C and the estimated secondary path $\tilde{C}$ has to be in the region of $-\pi/2$ to $\pi/2$ for the MSE to converge. This condition is equivalent to that the argument of $C(\omega)/\tilde{C}(\omega)$ has to exist in the region of $-\pi/2$ to $\pi/2$. Here, ω denotes the angular frequency. $C(\omega)$ and $\tilde{C}(\omega)$ denote the frequency responses of C and $\tilde{C}$, respectively.

When $K = 2$, we obtain

$$\frac{C(\omega)}{\tilde{C}(\omega)} = \frac{c_1 + c_2 e^{-j\omega}}{\tilde{c}_1 + \tilde{c}_2 e^{-j\omega}}. \tag{50}$$

We can easily calculate the real part of (50) as

$$\frac{c_1\tilde{c}_1 + c_2\tilde{c}_2 + (c_1\tilde{c}_2 + \tilde{c}_1 c_2)\cos\omega}{((\tilde{c}_1 + \tilde{c}_2)\cos\omega)^2 + (\tilde{c}_2 \sin\omega)^2}. \tag{51}$$

Here, the denominator of (51) is obviously positive. Therefore, the numerator of (51) should be positive for the argument of $C(\omega)/\tilde{C}(\omega)$ to exist in the region of $-\pi/2$ to $\pi/2$. That is, the necessary condition for stability is

$$c_1\tilde{c}_1 + c_2\tilde{c}_2 + (c_1\tilde{c}_2 + \tilde{c}_1 c_2)\cos\omega > 0. \tag{52}$$

Substituting $c_1 = c_2 = \tilde{c}_1 = 1$, that is, the conditions in this subsection, into (52), we obtain

$$(1 + \tilde{c}_2)(1 + \cos\omega) > 0. \tag{53}$$

Therefore, the condition for stability is

$$\tilde{c}_2 > -1. \tag{54}$$

It can be seen that the behavior in Fig. 10 is consistent with (54).

However, the 90° condition is valid for the infinitely small step size μ [41], [42]. This means that the 90° condition is not a sufficient condition but a necessary condition when the step size μ is not infinitely small. Therefore, the right-hand side of (54) does not agree with the value of $\tilde{c}_2$ for which the MSE diverges in Fig. 10.

The necessary and sufficient condition in the case where the step size is not infinitely small was derived in [41], [42], written by one of the authors of this paper. The condition is

$$\left|1 - \mu\|\boldsymbol{x}_h(n)\|^2 C(\omega)\tilde{C}^*(\omega)\right| < 1, \tag{55}$$

where $*$ denotes the complex conjugate.

Substituting $\mu = 0.1$, $\|\boldsymbol{x}_h(n)\| = 1$, and $c_1 = c_2 = \tilde{c}_1 = 1$, that is, the conditions in this subsection, into (55) and numerically calculating the lower bound of $\tilde{c}_2$ satisfying (55) for arbitrary ω in the region of $-\pi < \omega < \pi$, we obtain -0.7802 as the lower bound of $\tilde{c}_2$. This bound is exactly in agreement with the behavior in Fig. 10. This indicates that the theory derived in this paper and the theory derived in [41], [42] justify each other. The authors consider this justification to be significant.

5. Conclusions

We have analyzed the behaviors of active noise control with the FXLMS algorithm using a statistical-mechanical method. The obtained theory can treat not only a white reference signal but also a nonwhite reference signal. In addition, we have also shown that the theory can predict the behaviors in the case of an actual primary path by using the autocorrelations of the actual primary path. Furthermore, the effect of the estimation error of the secondary path was investigated using the derived theory. The theory quantitatively agrees with the results of computer simulations. The principal assumption used in this paper is that the tapped-delay lines of the primary path and the adaptive filter are sufficiently long.

Acknowledgments

One of the authors (Seiji Miyoshi) thanks Professor Tatsuya Uezu for helpful discussions. This work was supported by JSPS KAKENHI Grant Numbers JP24360152, JP24560280, JP15K00256, JP17K06449, and JP18K11483.

References

[1] P.A. Nelson and S.J. Elliott, Active Control of Sound, Academic, San Diego, CA, 1992.
[2] S.M. Kuo and D.R. Morgan, Active Noise Control Systems — Algorithms and DSP Implementations, Wiley, New York, 1996.
[3] S.M. Kuo and D.R. Morgan, "Active noise control: A tutorial review," Proc. IEEE, vol.87, no.6, pp.943–973, June 1999.
[4] Y. Kajikawa, W.-S. Gan, and S.M. Kuo, "Recent advances on active noise control: Open issues and innovative applications," APSIPA Trans. Signal Inf. Process., vol.1, e3, Aug. 2012.
[5] B. Widrow and M.E. Hoff, Jr., "Adaptive switching circuits," IRE WESCON Conv. Rec., Pt.4, pp.96–104, 1960.
[6] S. Haykin, Adaptive Filter Theory, Fourth ed., Prentice Hall, Upper Saddle River, NJ, 2002.
[7] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley, Hoboken, NJ, 2003.
[8] B. Widrow and S.D. Stearns, Adaptive Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1985.
[9] B. Widrow, J.M. McCool, M.G. Larimore, and C.R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," Proc. IEEE, vol.64, no.8, pp.1151–1162, Aug. 1976.
[10] J.E. Mazo, "On the independence theory of equalizer convergence," Bell Syst. Tech. J., vol.58, no.5, pp.963–993, 1979.
[11] W.A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique," Signal Process., vol.6, no.2, pp.113–133, 1984.
[12] E. Bjarnason, "Analysis of the Filtered-X LMS algorithm," IEEE Trans. Speech Audio Process., vol.3, no.6, pp.504–514, Nov. 1995.
[13] I.T. Ardekani and W. Abdulla, "Theoretical convergence analysis of FxLMS algorithm," Signal Process., vol.90, no.12, pp.3046–3055, 2010.
[14] I.T. Ardekani and W. Abdulla, "On the convergence of real-time active noise control systems," Signal Process., vol.91, no.5, pp.1262–1274, 2011.
[15] S. Chan and Y. Chu, "Performance analysis and design of FxLMS algorithm in broadband ANC system with online secondary-path modeling," IEEE Trans. Audio Speech Lang. Process., vol.20, no.3, pp.982–993, March 2012.
[16] M.H. Costa, J.C.M. Bermudez, and N.J. Bershad, "Stochastic analysis of the Filtered-x LMS algorithm in systems with nonlinear secondary paths," IEEE Trans. Signal Process., vol.50, no.4, pp.1327–1342, 2002.
[17] O.J. Tobias, J.C.M. Bermudez, and N.J. Bershad, "Mean weight behavior of the Filtered-X LMS algorithm," IEEE Trans. Signal Process., vol.48, no.4, pp.1061–1075, April 2000.
[18] O.J. Tobias, J.C.M. Bermudez, R. Seara, and N. Bershad, "An improved model for the second moment behavior of the Filtered-X LMS algorithm," Proc. IEEE Adaptive Syst. Signal Process., Commun., Contr. Symp., pp.337–341, Lake Louise, AB, Canada, Oct. 2000.
[19] L.S. Resende and J.C.M. Bermudez, "An efficient model for the convergence behavior of the FXLMS algorithm with Gaussian inputs," Proc. IEEE/SP 13th Workshop on Statistical Signal Processing, pp.97–102, 2005.
[20] H.J. Butterweck, "The independence assumption: A dispensable tool in adaptive filter theory," Signal Process., vol.57, no.3, pp.305–310, 1997.
[21] S. Miyagi and H. Sakai, "Mean-square performance of the filtered-reference/filtered-error LMS algorithm," IEEE Trans. Circuits Syst. I, Reg. Papers, vol.52, no.11, pp.2454–2463, Nov. 2005.
[22] L. Vicente and E. Masgrau, "Novel FxLMS convergence condition with deterministic reference," IEEE Trans. Signal Process., vol.54, no.10, pp.3768–3774, Oct. 2006.
[23] Y. Hinamoto and H. Sakai, "Analysis of the Filtered-X LMS algorithm and a related new algorithm for active control of multitonal noise," IEEE Trans. Audio, Speech, Language Process., vol.14, no.1, pp.123–130, Jan. 2006.
[24] Y. Hinamoto and H. Sakai, "A Filtered-X LMS algorithm for sinusoidal reference signals — Effects of frequency mismatch," IEEE Signal Process. Lett., vol.14, no.4, pp.259–262, April 2007.
[25] H. Nishimori, Statistical Physics of Spin Glasses and Information Processing: An Introduction, Oxford University Press, New York, 2001.
[26] M. Okada, "Notions of associative memory and sparse coding," Neural Networks, vol.9, no.8, pp.1429–1458, 1996.
[27] S. Miyoshi and M. Okada, "Storage capacity diverges with synaptic efficiency in an associative memory model with synaptic delay and pruning," IEEE Trans. Neural Netw., vol.15, no.5, pp.1215–1227, 2004.
[28] Y. Kabashima and D. Saad, "Statistical mechanics of error correcting codes," Europhys. Lett., vol.45, no.1, pp.97–103, 1999.
[29] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," IEEE Trans. Inf. Theory, vol.48, no.11, pp.2888–2910, 2002.
[30] K. Tanaka, "Statistical-mechanical approach to image processing," J. Phys. A, vol.35, no.37, pp.R81–R150, 2002.
[31] A. Engel and C.V. Broeck, Statistical Mechanics of Learning, Cambridge University Press, Cambridge, 2001.
[32] D. Saad, ed., On-Line Learning in Neural Networks, Cambridge University Press, Cambridge, 1998.
[33] S. Miyoshi, K. Hara, and M. Okada, "Analysis of ensemble learning using simple perceptrons based on online learning theory," Phys. Rev. E, vol.71, no.3, 036116, 2005.
[34] S. Miyoshi and Y. Kajikawa, "Statistical-mechanics approach to the Filtered-X LMS algorithm," Electron. Lett., vol.47, no.17, pp.997–999, Aug. 2011.
[35] S. Miyoshi and Y. Kajikawa, "Theoretical discussion of the Filtered-X LMS algorithm based on statistical mechanical analysis," Proc. IEEE Statistical Signal Processing Workshop (SSP2012), pp.341–344, 2012.
[36] S. Miyoshi and Y. Kajikawa, "Statistical-mechanical analysis of the FXLMS algorithm with nonwhite reference signals," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2013), pp.5652–5656, May 2013.
[37] S. Miyoshi and Y. Kajikawa, "Statistical-mechanical analysis of the FXLMS algorithm with actual primary path," Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP2015), pp.3502–3506, April 2015.
[38] L. Ljung, "Analysis of recursive stochastic algorithms," IEEE Trans. Autom. Control, vol.22, no.4, pp.551–575, 1977.
[39] H.J. Kushner and D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer-Verlag, New York, 1978.
[40] S. Snyder and C. Hansen, "The effect of transfer function estimation errors on the Filtered-x LMS algorithm," IEEE Trans. Signal Process., vol.42, no.4, pp.950–953, April 1994.
[41] J. Yabuki, Y. Kajikawa, and Y. Nomura, "A derivation of the stable condition for the filtered-x LMS algorithm in the case where $\hat{C}$ filter has modeling error," IEICE Trans. Fundamentals (Japanese Edition), vol.J80-A, no.11, pp.1868–1876, Nov. 1997.
[42] Y. Kajikawa, J. Yabuki, and Y. Nomura, "Stable condition considering modeling error in the filtered-x LMS algorithm," Proc. IEEE Int. Symp. Circuits and Systems (ISCAS2000), pp.642–645, May 2000.

Appendix A: Justification of (17)

From (3) and (7), we obtain

$$\boldsymbol{h}(n)^T \boldsymbol{x}_h(n) = \sum_{i=1}^{N_h} h_i(n) x(n-i+1). \tag{A·1}$$

If $j \geq 0$, from (15) and (16), we obtain

$$\begin{aligned}
&\boldsymbol{k}_j(n)^T \boldsymbol{x}_h(n-j) \\
&= \sum_{i=1}^{N_h} k_{j,i}(n) x(n-j-i+1) && \text{(A·2)} \\
&= \sum_{i=1}^{N_h} h_{\mathrm{mod}(i+j-1,N_h)+1}(n) x(n-j-i+1) && \text{(A·3)} \\
&= \sum_{i=1}^{N_h-j} h_{i+j}(n) x(n-j-i+1) \\
&\quad + \sum_{i=N_h-j+1}^{N_h} h_{i+j-N_h}(n) x(n-j-i+1) && \text{(A·4)} \\
&= \sum_{i=1+j}^{N_h} h_i(n) x(n-i+1) \\
&\quad + \sum_{i=1}^{j} h_i(n) x(n-i+1-N_h). && \text{(A·5)}
\end{aligned}$$

$$\begin{aligned}
&\therefore \frac{\boldsymbol{k}_j(n)^T \boldsymbol{x}_h(n-j)}{\boldsymbol{h}(n)^T \boldsymbol{x}_h(n)} \\
&= \left(\sum_{i=1}^{N_h} h_i(n) x(n-i+1) - \sum_{i=1}^{j} h_i(n) x(n-i+1) \right. \\
&\quad \left. + \sum_{i=1}^{j} h_i(n) x(n-i+1-N_h) \right) \Bigg/ \\
&\qquad \sum_{i=1}^{N_h} h_i(n) x(n-i+1) && \text{(A·6)} \\
&= 1 + \frac{\sum_{i=1}^{j} h_i(n) \left(x(n-i+1-N_h) - x(n-i+1) \right)}{\sum_{i=1}^{N_h} h_i(n) x(n-i+1)} && \text{(A·7)}
\end{aligned}$$

The denominator and numerator of the second term of the right-hand side of (A·7) is $O(1)$ and $O(N_h^{-\frac{1}{2}})$, respectively. Therefore, we obtain

$$\lim_{N_h \to \infty} \frac{\boldsymbol{k}_j(n)^T \boldsymbol{x}_h(n-j)}{\boldsymbol{h}(n)^T \boldsymbol{x}_h(n)} = 1. \tag{A·8}$$

For $j < 0$, the same result is derived in the same manner.

Appendix B: Derivation of (22)

Since the long-filter condition $N_p, N_h \to \infty$ is assumed, from (5), (8), (15)–(21), and the law of large numbers, we obtain

$$\begin{aligned}
&\langle d(n-j) u(n) \rangle \\
&= \left\langle \boldsymbol{p}^T \boldsymbol{x}_p(n-j) \boldsymbol{h}(n)^T \boldsymbol{x}_h(n) \right\rangle && \text{(A·9)} \\
&= \left\langle \sum_{i=1}^{N_p} p_i x(n-j-i+1) \sum_{i'=1}^{N_h} h_{i'}(n) x(n-i'+1) \right\rangle && \text{(A·10)} \\
&= \sum_{i=1}^{N_p} \sum_{i'=1}^{N_h} p_i h_{i'}(n) \langle x(n-j-i+1) x(n-i'+1) \rangle && \text{(A·11)} \\
&= \sum_{i=1}^{N_p} \sum_{i'=1}^{N_h} p_i k_{i'-i,i}(n) \frac{r_{i'-i-j}}{N_h} && \text{(A·12)} \\
&= \sum_{k=-M}^{M} \left(\min(N_p, N_h) - k \right) R_k(n) \frac{r_{k-j}}{N_h} && \text{(A·13)} \\
&\quad (\because k \equiv i' - i, R_{k'}(n) = 0 \text{ when } |k'| > M) \\
&\xrightarrow{N_p, N_h \to \infty} \sum_{k=-M}^{M} \bar{a} R_k(n) r_{k-j}. && \text{(A·14)}
\end{aligned}$$

When transforming (A·10) into (A·11), we assumed that the correlation between $\boldsymbol{h}(n)$ and $\boldsymbol{x}_h(n)$ is small [16]–[18].

Appendix C: Derivation of (24)

Since the long-filter condition $N_p, N_h \to \infty$ is assumed, from (2), (5), (18), and the law of large numbers, we obtain

$$\begin{aligned}
&\langle d(n-j) d(n) \rangle \\
&= \left\langle \boldsymbol{p}^T \boldsymbol{x}_p(n-j) \boldsymbol{p}^T \boldsymbol{x}_p(n) \right\rangle && \text{(A·15)} \\
&= \left\langle \sum_{i=1}^{N_p} p_i x(n-j-i+1) \sum_{i'=1}^{N_p} p_{i'} x(n-i'+1) \right\rangle && \text{(A·16)} \\
&= \sum_{i=1}^{N_p} \sum_{i'=1}^{N_p} p_i p_{i'} \langle x(n-j-i+1) x(n-i'+1) \rangle && \text{(A·17)} \\
&= \sum_{i=1}^{N_p} \sum_{i'=1}^{N_p} p_i p_{i'} \frac{r_{i'-i-j}}{N_h} && \text{(A·18)} \\
&= \sum_{k=-L}^{L} \left(N_p - k \right) \kappa_k \frac{r_{k-j}}{N_h} && \text{(A·19)} \\
&\quad (\because k \equiv i' - i, \kappa_{k'} = 0 \text{ when } |k'| > L) \\
&= \sum_{k=-L}^{L} \left(a - \frac{k}{N_h} \right) \kappa_k r_{k-j} && \text{(A·20)} \\
&\xrightarrow{N_h \to \infty} \sum_{k=-L}^{L} a \kappa_k r_{k-j}. && \text{(A·21)}
\end{aligned}$$

Appendix D: Outline of Derivation of (36)

Multiplying (13) by (26), we obtain

$$\begin{aligned} &N_h Q_j(n+1) \\ &= N_h Q_j(n) \\ &+ \mu e(n) \sum_{k'=1}^{K} \tilde{c}_{k'} \boldsymbol{x}_h(n-k'+1)^T \boldsymbol{k}_j(n) \\ &+ \mu e(n) \sum_{k'=1}^{K} \tilde{c}_{k'} \boldsymbol{x}_h(n-k'+1-j)^T \boldsymbol{h}(n) \\ &+ \mu^2 e^2(n) \sum_{k'=1}^{K} \sum_{k''=1}^{K} \tilde{c}_{k'} \tilde{c}_{k''} \\ &\qquad \times \boldsymbol{x}_h(n-k'+1)^T \boldsymbol{x}_h(n-k''+1-j). \end{aligned} \tag{A·22}$$

From (13), we obtain

$$\boldsymbol{h}(n) = \boldsymbol{h}(n+1) - \mu e(n) \sum_{k=1}^{K} \tilde{c}_k \boldsymbol{x}_h(n-k+1). \tag{A·23}$$

Using $\Delta(n)$, which is defined by $\Delta(n) = \mu e(n) \sum_{k=1}^{K} \tilde{c}_k \boldsymbol{x}_h(n-k+1)$, we can write

$$\boldsymbol{h}(n) = \begin{cases} \boldsymbol{h}(n+j) - \sum_{k=1}^{j} \Delta(n+j-k), & j \geq 0 \\ \boldsymbol{h}(n+j) + \sum_{k=1}^{-j} \Delta(n-k), & j < 0. \end{cases} \tag{A·24}$$

Using the sign function and step function defined by (38), we can represent (A·24) by a single formula,

$$\boldsymbol{h}(n) = \boldsymbol{h}(n+j) - \mathrm{sgn}(j) \sum_{k=1}^{|j|} \Delta(n + \Theta(j)j - k). \tag{A·25}$$

From (17) and (A·25), we rewrite $\boldsymbol{x}_h(n-k'+1)^T \boldsymbol{k}_j(n)$ in the second term on the right-hand side of (A·22) as

$$\boldsymbol{x}_h(n-k'+1)^T \boldsymbol{k}_j(n) = \boldsymbol{x}_h(n-k'+1+j)^T \boldsymbol{h}(n) \tag{A·26}$$

$$\begin{aligned} &= \boldsymbol{x}_h(n+j+1-k')^T \Bigg(\boldsymbol{h}(n+j+1-k') \\ &\quad - \mathrm{sgn}(j+1-k') \\ &\qquad \times \sum_{k=1}^{|j+1-k'|} \Delta(n + \Theta(j+1-k')(j+1-k') - k) \Bigg). \end{aligned} \tag{A·27}$$

From (8), (10), and (11), we can calculate (A·27) as

$$\begin{aligned} &u(n+\gamma) - \mu \mathrm{sgn}(\gamma) \sum_{k''=1}^{\gamma} \Bigg(d(n + \Theta(\gamma)\gamma - k'') \\ &\quad - \sum_{k'''=1}^{K} c_{k'''} u(n + \Theta(\gamma)\gamma - k'' - k''' + 1) \\ &\quad + \xi(n + \Theta(\gamma)\gamma - k'') \Bigg) \sum_{k=1}^{K} \tilde{c}_i r_{\Theta(\gamma)\gamma - k'' - i + k' - j}, \end{aligned} \tag{A·28}$$

where $\gamma = j + 1 - k'$. We set $\tilde{c}_k = 0$ when $k < 1$ or $k > K$ in (A·28).

Using (8)–(11) and (A·28), we can calculate the mean of the second term on the right-hand side of (A·22) as

$$\begin{aligned} &\left\langle \mu e(n) \sum_{k'=1}^{K} \tilde{c}_{k'} \boldsymbol{x}_h(n-k'+1)^T \boldsymbol{k}_j(n) \right\rangle \\ &= \mu \sum_{k'=1}^{K} \tilde{c}_{k'} \Bigg\{ \sum_{i=-M}^{M} \left(\bar{a} R_i r_{i-\gamma} - \sum_{k=1}^{K} c_k Q_i r_{i-k'+k+j} \right) \\ &\qquad - \mu \mathrm{sgn}(\gamma) \sum_{k''=1}^{|\gamma|} \Bigg[\delta_{\alpha,0} \sigma_\xi^2 + a \sum_{l=-L}^{L} \kappa_l r_{l+\alpha} \\ &\qquad - \sum_{k=1}^{K} c_k \sum_{i=-M}^{M} \Big(\bar{a} R_i r_{i+k-1-\alpha} + \bar{a} R_i r_{i+k-1+\alpha} \\ &\qquad\qquad - \sum_{k'''=1}^{K} c_{k'''} Q_i r_{i-k+k'''-\alpha} \Big) \Bigg] \\ &\qquad\qquad \times \sum_{i=1}^{K} \tilde{c}_i r_{k'-i-j+\alpha} \Bigg\}, \end{aligned} \tag{A·29}$$

where $\alpha = \Theta(\gamma)\gamma - k''$.

In the same manner as that used to obtain (A·29), we can calculate the means of the third and fourth terms in (A·22). Finally, we obtain the differential equation for Q as (36).

Seiji Miyoshi received his B.Eng. and M.Eng. degrees in electrical engineering from Kyoto University, Japan, in 1986 and 1988, respectively, and his Ph.D degree in system science and engineering from Kanazawa University, Japan, in 1998. He worked with the Space Development Division of NEC Corporation from 1988 to 1994. He joined the Department of Electronic Engineering of Kobe City College of Technology in 1994. He joined the Department of Electrical and Electronic Engineering, the Faculty of Engineering Science of Kansai University in 2008, where he is a professor. His research interests include statistical-mechanical informatics, image processing, signal processing, and learning theory. Dr. Miyoshi is a senior member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Electrical and Electronics Engineers (IEEE), and a member of the Physical Society of Japan (JPS), and the Japan Neural Network Society (JNNS). He served as an associate editor for IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences from 2013 to 2017 and an awards committee member of IEEE Kansai section from 2011 to 2014. He received the Papers of Editors' Choice of Journal of the Physical Society of Japan in 2011 and Excellent Paper Award of ICONIP2016.

Yoshinobu Kajikawa received his B.Eng. and M.Eng. degrees in electrical engineering from Kansai University, Osaka, Japan, and his D.E. degree in communication engineering from Osaka University in 1991, 1993, and 1997, respectively. He joined Fujitsu Ltd., Kawasaki, Japan, in 1993, where he was engaged in research on active noise control. In 1994, he joined Kansai University, where he is now a professor. His current research interests lie in the area of signal processing for audio and acoustic systems. Dr. Kajikawa is a senior member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Electrical and Electronics Engineers (IEEE), and a member of the European Association for Signal Processing (EURASIP), the Asia-Pacific Signal and Information Processing Association (APSIPA), the Acoustical Society of Japan (ASJ), and the Acoustical Society of America (ASA). He served as an associate editor for IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences from 2006 to 2009. He is currently serving as an associate editor for the IET Signal Processing and as a vice president in APSIPA. He is the author or co-author of more than 200 articles published in journals and conference proceedings and has been granted eight patents. He received the 2012 Sato Prize Paper Award from the ASJ, the best paper award in APCCAS 2014, and the Sadaoki Furui Prize Paper Award from the APSIPA.