Statistical-Mechanical Analysis of the Second-Order Adaptive Volterra Filter

Kimiko Motonaka*, Takashi Katsube†, Yoshinobu Kajikawa*, and Seiji Miyoshi*
* Kansai University, Osaka, Japan
E-mail: motonaka@kansai-u.ac.jp Tel: +86-6368-1121
† KYOCERA Communication Systems Co., Ltd., Japan

***Abstract*—The Volterra filter is a digital filter that can describe nonlinearity. In this paper, we analyze the dynamic behaviors of an adaptive signal-processing system based on the Volterra filter by a statistical-mechanical method. On the basis of the self-averaging when the tapped delay line is assumed to be infinitely long, we derive simultaneous differential equations in a deterministic and closed form that describe the behaviors of the macroscopic variables and obtain the exact solution by solving them analytically. In addition, the validity of the derived theory is confirmed by comparison with numerical simulations.**

I. Introduction

Adaptive signal-processing techniques have been used in various fields, such as information communication and acoustic systems[1], [2]. There have been many studies on techniques using a linear digital filter and also on theoretical analysis [3], [4], [5]. As an example, a linear digital filter for active noise control [6], [7], [8], [9] has been subjected to statistical-mechanical analysis [10]. On the other hand, in an actual environment, the target system is often nonlinear, thus, a digital filter that can describe the nonlinearity is desirable. The Volterra filter is one such digital filter that can describe nonlinearity [11]. The Volterra filter can effectively describe weak nonlinearity with no hysteresis and has been applied to the modeling of loudspeaker systems and other applications. However, there has been insufficient theoretical analysis of the Volterra filter. Therefore, in this paper, we analyze the dynamic behaviors of an adaptive signal-processing system based on the Volterra filter by a statistical-mechanical method. Such theoretical analysis is important for obtaining deep insight into the behavior of the system.

While many update algorithms have been proposed for the adaptive Volterra filter, we analyze the case of using the least-mean-square (LMS) algorithm [1], [2], [12]. In this paper, on the basis of self-averaging [10], when the tapped delay line is assumed to be infinitely long, we derive simultaneous differential equations in a deterministic and closed form that describe the behaviors of the macroscopic variables and obtain the exact solution by solving them analytically. Then, we verify the behavior of the Volterra filter while changing the step size and background noise. Finally, we confirm that the analytically obtained dynamic behavior of the mean square error MSE (learning curves) is in good agreement results of numerical simulations.

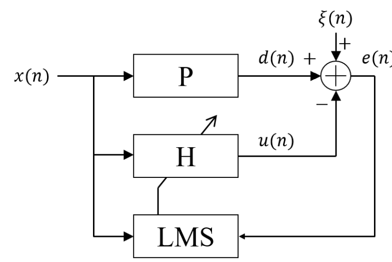

Fig. 1. Block diagram of the adaptive system.

II. Volterra Filter

The Volterra filter is a digital filter that can describe nonlinearity and uses the Volterra kernel of the Volterra series expansion as the digital filter. The discrete Volterra series expansion up to the Lth Volterra kernel is defined as

$$\begin{aligned} y(n) = & \sum_{k_1=0}^{\infty} h(k_1)x(n-k_1) \\ & + \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} h(k_1,k_2)x(n-k_1)x(n-k_2) \\ & + ... \\ & + \sum_{k_1=0}^{\infty} ... \sum_{k_L=0}^{\infty} h(k_1,...,k_L)\prod_{i=1}^{L} x(n-k_i). \end{aligned} \tag{1}$$

Here, $x(n)$ and $y(n)$ are the input signal and output signal of time step n, respectively. $h(k_1,...,k_l)$ is the l-th Volterra kernel. In the adaptive Volterra filter, this Volterra kernel $h(k_1,...,k_l)$ is updated. Note that, in the case of $L = 1$, it is equivalent to a linear filter.

III. Analytical Model

The Volterra filter applied to adaptive signal processing is called the adaptive Volterra filter. Various methods, which are the same as those used with a normal linear adaptive filter, for example, the gradient method and the recursive least-squares (RLS) method, can be used to update the adaptive Volterra filter, and we analyze the case when the LMS algorithm, which is one of the gradient methods, is used for updating. Figure 1 shows a block diagram of the adaptive system. In Fig. 1, P denotes the unknown system and H denotes the adaptive filter. P and H are constructed from the second-order Volterra kernels, $\boldsymbol{P} = \{p(k_1, k_2)\}$, $\boldsymbol{H}(n) = \{h(k_1, k_2; n)\}$, $k_1, k_2 = 0, ..., N-1$. Each element of upper triangular part of the matrix $\boldsymbol{P}$, that is, $p(k1, k2)$, $k1 \leq k2$, is generated independently from a distribution with a mean of zero and a variance of one. In addition, each element $p(k2, k1)$, $k1 \leq k2$ is equal to $p(k1, k2)$. This means P is a symmetric matrix. The initial matrix $\boldsymbol{H}(0)$ is set to the zero matrix. The input signal $x(n)$ is generated independently from a distribution with a mean of zero and a variance of $1/N$. Here, the tap input vector $\boldsymbol{x}(n)$ in time step n is

$$\boldsymbol{x}(n) = [x(n), x(n-1), ..., x(n-N+1)]^T. \tag{2}$$

The outputs $d(n)$ and $u(n)$ of P and H, respectively, in time step n are

$$\begin{aligned} d(n) &= \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) x(n-k_1) x(n-k_2) \\ &= \boldsymbol{x}(n)^T \boldsymbol{P} \boldsymbol{x}(n), \end{aligned} \tag{3}$$

$$\begin{aligned} u(n) &= \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} h(k_1, k_2; n) x(n-k_1) x(n-k_2) \\ &= \boldsymbol{x}(n)^T \boldsymbol{H}(n) \boldsymbol{x}(n). \end{aligned} \tag{4}$$

The error signal $e(n)$ is obtained by subtracting the output signal of the adaptive Volterra filter $u(n)$ from the output signal of the unknown system $d(n)$ and adding a background noise $\xi(n)$:

$$e(n) = d(n) - u(n) + \xi(n). \tag{5}$$

Here, the background noise $\xi(n)$ is generated independently from a distribution with a mean of zero and a variance of σ_ξ^2. The update formula of the adaptive Volterra filter using the LMS algorithm is

$$h(k_1, k_2; n+1) = h(k_1, k_2; n) + \mu e(n) x(n-k_1) x(n-k_2), \tag{6}$$

where μ denotes the step-size parameter.

IV. Theory

In this section, we describe a theoretical analysis of the behaviors of the adaptive Volterra filter by a statistical-mechanical method. The MSE of the model used can be calculated using Eq. (5) as follows:

$$\left\langle e^2(n) \right\rangle = \left\langle (d(n) - u(n) + \xi(n))^2 \right\rangle \tag{7}$$

$$= \left\langle d^2(n) \right\rangle + \left\langle u^2(n) \right\rangle - 2 \left\langle d(n) u(n) \right\rangle + \sigma_\xi^2. \tag{8}$$

Note that the background noise $\xi(n)$ is independent of the other stochastic variables and its variance σ_ξ^2 is used. The detail of the calculation is omitted, but each term on the right side of Eq. (8) can be calculated using Eqs. (3) and (4) as follows:

$$\left\langle d^2(n) \right\rangle = \frac{2}{N^2} \sum_{k_1=0}^{N-1} \sum_{\substack{k_2=0 \\ k_2 \neq k_1}}^{N-1} p^2(k_1, k_2), \tag{9}$$

$$\left\langle u^2(n) \right\rangle = \frac{2}{N^2} \sum_{k_1=0}^{N-1} \sum_{\substack{k_2=0 \\ k_2 \neq k_1}}^{N-1} h^2(k_1, k_2; n), \tag{10}$$

$$\left\langle d(n) u(n) \right\rangle = \frac{2}{N^2} \sum_{k_1=0}^{N-1} \sum_{\substack{k_2=0 \\ k_2 \neq k_1}}^{N-1} p(k_1, k_2) h(k_1, k_2; n). \tag{11}$$

Here, we assume that there is little correlation between $\boldsymbol{x}(n)$ and $\boldsymbol{H}(n)$ [3], [4], [5]. In addition, it is later assumed that $N \to \infty$, and the terms that can be ignored when $N \to \infty$ are omitted here. Next, we introduce the macroscopic variables $R(n)$ and $Q(n)$, defined as follows:

$$R(n) = \frac{1}{N^2} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) h(k_1, k_2; n), \tag{12}$$

$$Q(n) = \frac{1}{N^2} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} h^2(k_1, k_2; n). \tag{13}$$

From Eqs. (8)-(13),

$$\left\langle e^2(n) \right\rangle = \left\langle d^2(n) \right\rangle + 2Q(n) - 4R(n) + \sigma_\xi^2. \tag{14}$$

This equation shows that the MSE is a function of the macroscopic variables $R(n)$ and $Q(n)$.

Now, we derive the simultaneous differential equations that describe the dynamic behavior of the MSE. First, we derive the differential equation for $R(n)$. Multiplying both sides of Eq. (6) by $p(k_1, k_2)$ and summing it over k_1 and k_2, we obtain

$$\begin{aligned} &\sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) h(k_1, k_2; n+1) \\ &\quad = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) h(k_1, k_2; n) \\ &\quad + \mu e(n) \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} p(k_1, k_2) x(n-k_1) x(n-k_2). \end{aligned} \tag{15}$$

From Eqs. (3) and (12),

$$N^2 R(n+1) = N^2 R(n) + \mu e(n) d(n). \tag{16}$$

Note that the first terms on both sides of Eq. (16) are $O(N^2)$ but the second term on the right side is $O(1)$. Thus, to change $R(n)$ by $O(1)$, $O(N^2)$ updates are needed. Therefore, we use the value t, which is n normalized by N^2, as the time scale. By updating $N^2 dt$ times in an infinitely small time dt, we can obtain $N^2 dt$ equations similar to Eq. (16). Summing all equations, we obtain

$$N^2R(n+N^2dt) = N^2R(n)+\mu\left(\sum_{i=0}^{N^2dt-1} e(n+i)d(n+i)\right). \quad (17)$$

Hereafter, we assume that $N \to \infty$. Then, the second term on the right side of Eq. (17) can be replaced by its mean[1] as follows:

$$N^2R(n+N^2dt) = N^2R(n)+N^2dt\mu\left\langle e(n)d(n)\right\rangle. \quad (18)$$

Defining the change in R after updating N^2dt times as dR, we obtain

$$\frac{dR(t)}{dt} = \mu\left\langle e(n)d(n)\right\rangle. \quad (19)$$

From Eqs. (5), (9), (11), (12), and (19), the differential equation for R can be obtained as

$$\frac{dR(t)}{dt} = 2\mu(1-R(t)). \quad (20)$$

Second, we derive the differential equation for Q. Squaring both sides of Eq. (6), and summing over k_1 and k_2, we obtain

$$\begin{aligned}\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} & h^2(k_1,k_2;n+1)\\ &= \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} h^2(k_1,k_2;n)\\ &\quad + 2\mu e(n)\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} h(k_1,k_2;n)x(n-k_1)x(n-k_2)\\ &\quad + \mu^2e^2(n)\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} x^2(n-k_1)x^2(n-k_2). \quad (21)\end{aligned}$$

This can be rewritten as follows using Eqs. (4) and (13):

$$\begin{aligned}N^2Q(n+1) &= N^2Q(n)+2\mu e(n)u(n)\\ &\quad + \mu^2e^2(n)\sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} x^2(n-k_1)x^2(n-k_2). \quad (22)\end{aligned}$$

Similarly to the case of R, by updating N^2dt times in an infinitely small time dt and summing all equations, we obtain

$$\begin{aligned}N^2Q(n+N^2dt) = N^2Q(n) &+ 2N^2dt\mu\left\langle e(n)u(n)\right\rangle\\ &+N^2dt\mu^2\left\langle e^2(n)\right\rangle. \quad (23)\end{aligned}$$

From Eqs. (5), (9), and (11)-(14), we obtain

$$\begin{aligned}\frac{dQ(t)}{dt} &= 2\mu\left\langle e(n)u(n)\right\rangle+\mu^2\left\langle e^2(n)\right\rangle \quad (24)\\ &= 4\mu(R(t)-Q(t))\\ &\quad +\mu^2(2+2Q(t)-4R(t)+\sigma_\xi^2). \quad (25)\end{aligned}$$

[1]This characteristic is called self-averaging in statistical mechanics[10].

The derived differential equations for R and Q (Eqs. (20) and (25), respectively) can be solved analytically, and we obtain

$$R(t) = 1-e^{-2\mu t} \quad (26)$$

$$Q(t) = 1+\left(1+\frac{\mu\sigma_\xi^2}{2(\mu-2)}\right)e^{2\mu(\mu-2)t}-\frac{\mu\sigma_\xi^2}{2(\mu-2)}-2e^{-2\mu t}. \quad (27)$$

Substituting these equations in Eq. (14), we can obtain the exact solution of the MSE as

$$\left\langle e^2(t)\right\rangle = \left(2+\frac{\mu\sigma_\xi^2}{\mu-2}\right)e^{2\mu(\mu-2)t}-\frac{2\sigma_\xi^2}{\mu-2}. \quad (28)$$

Here, we used the fact that $\left\langle d^2(n)\right\rangle = 1$ in Eq. (14) when $N \to \infty$.

V. Results and Discussion

From the exact solution, Eq. (28), found in the previous section, we can obtain deep insight into the behavior of the MSE. For example, a necessary and sufficient condition for convergence of the MSE is $0 < \mu < 2$. In this case, the steady-state value of the MSE is $2\sigma_\xi^2/(2-\mu)$. Additionally, if there is no background noise, the MSE is

$$\left\langle e^2(t)\right\rangle = e^{2\mu(\mu-2)t} = e^{-2(1-(\mu-1)^2)t}, \quad (29)$$

thus, the MSE takes a minimum value at $\mu = 1.0$ regardless of the value of t.

Figures 2 and 3 show the learning curves with $\sigma_\xi^2 = 0$, i.e., the background noise is zero. In the figures, the solid lines denote the theoretical results and the symbols denote the values in numerical simulations. In the numerical simulations, the tap length was set to $N = 100, 200$ and the mean values of five trials are plotted. The open symbols show the results for $N = 100$ and the filled symbols show the results for $N = 200$. As shown in Figs. 2 and 3, the MSE becomes minimum when $\mu = 1.0$ regardless of the value of t. In the numerical simulations, a systematic error occurs as μ increases. This is considered to be due to the finite-size effects [10] of the tap length N. Certainly, as shown in Figs. 2 and 3, the simulation results approach the theoretical results as the tap length N increases.

Figure 4 shows the learning curves with $\sigma_\xi^2 > 0$, i.e., background noise exists. In the figure, the solid lines denote the theoretical results and the symbols denote the values in the numerical simulations. In the numerical simulations, the tap length was set to $N = 100$ and the mean values of 100 trials are plotted. From Fig. 4, we can confirm that the rate of decrease of the MSE in the early stage is greatest for $\mu = 1$, but after sufficient time has elapsed, the MSE becomes smaller as μ decreases, thus the learning curves intersect.

From the above results, it was confirmed that the exact solution of the MSE derived in this paper is in good agreement with the results of the numerical simulations. The fact that the exact solution of the learning curves could be derived is

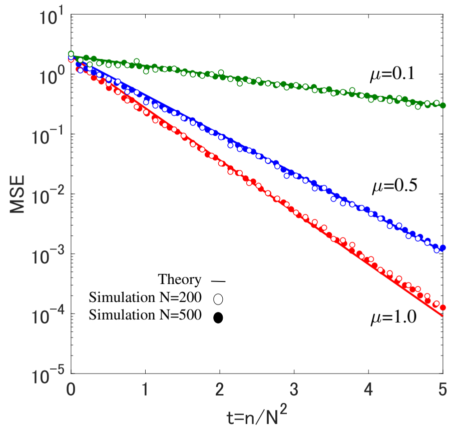

Fig. 2. Learning curves ($\sigma_\xi^2 = 0, \mu = 0.1, 0.5, 1.0$).

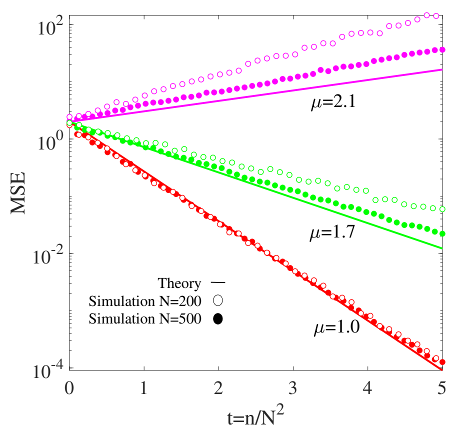

Fig. 3. Learning curves ($\sigma_\xi^2 = 0$, $\mu = 1.0, 1.5, 1.7, 2.1$).

significant from the both standpoints of obtaining deep insight into the Volterra filtering and the application of it. In the statistical-mechanical analysis of the linear FIR filter, the value t, which is the number of updates n normalized by the tap length N, is used as the time scale [6], [7], [8]. However, in the analysis in this paper, it was necessary to normalize the number of updates n by the tap length N^2 as the time scale t. This fact shows the essential slowness of the adaptation of the Volterra filter.

VI. Conclusions

In this paper, we analyzed the dynamic behaviors of an adaptive signal processing system based on the Volterra filter by a statistical-mechanical method. On the basis of self-averaging when the tapped delay line is assumed to be infinitely long, we derived simultaneous differential equations in a deterministic and closed form that describe the behaviors of the macroscopic variables, and obtained the exact solution by solving them analytically. In addition, the validity of the

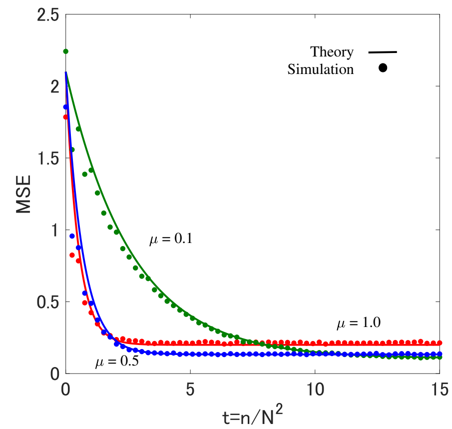

Fig. 4. Learning curves ($\sigma_\xi^2 = 0.1$, $\mu = 0.1, 0.5, 1.0$).

derived theory was confirmed by comparison with numerical simulations. Analysis of the case of $L \geq 3$ is one of our future works.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers 17K06449, 18K13782, and 18K11483, and the Kansai University Fund for Supporting Young Scholars, 2017.

References

[1] S. Haykin, *Adaptive Filter Theory*, 4^{th} ed., Prentice Hall, Upper Saddle River, NJ, 2002.

[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, Hoboken, NJ, 2003.

[3] 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. 6, pp. 1327–1342 (2002).

[4] 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 (2000).

[5] O. J. Tobias, J. C. M. Bermudez, R. Seara, and N. J. Bershad, "An improved model for the second moment of the filtered-X LMS algorithm," *Proc. of IEEE Adaptive Systems for Signal Processing, Communications, and Control Symp.*, 2000, pp. 337–341.

[6] S. Miyoshi and Y. Kajikawa, "Statistical-mechanical analysis of the FXLMS algorithm with nonwhite reference signals," *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 5652–5656.

[7] S. Miyoshi and Y. Kajikawa, "Statistical-mechanical analysis of the FXLMS algorithm with actual primary path," *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 3502–3506.

[8] S. Miyoshi and Y. Kajikawa, "Statistical-mechanics approach to the Filtered-X LMS algorithm," *Electron. Lett.*, vol. 47, no. 17, pp. 997–999, 2011.

[9] N. Ishibushi, Y. Kajikawa, and S. Miyoshi, "Statistical-mechanical analysis of LMS algorithm for time-varying unknown system," *J. Phys, Soc. Jpn.*, vol. 86, no. 2, 024803, 2017.

[10] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, New York, 2001.

[11] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, vol. 8, no. 3, pp. 10–26, 1991.

[12] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits", IRE WESCON *Conv. Rec.*, Pt. 4, pp. 96–104, 1960.