A Learning Method for System Identification

JIN-ICHI NAGUMO, MEMBER, IEEE, AND ATSUHIKO NODA

Abstract—A method for system identification is proposed which is based on the error-correcting training procedure in learning machines, and is referred to as "learning identification."

This learning identification is nondisturbing, is applicable to cases where the input signal is random and nonstationary, and can be completed within a short time, so that it may be used to identify linear quasi-time-invariant systems in which some parameters vary slowly in comparison with the time required for identification. This merit also makes it possible to eliminate noise disturbances by means of the moving average method.

Computer simulation of the learning identification was carried out and the times required for identification were obtained for various cases.

Some modifications of the learning identification were also investigated together with their computer simulations.

I. Introduction

A LINEAR time-invariant system can be identified in various ways. Some of these methods use particular input signals, such as sinusoid, step function, or impulse. These methods of system identification have the disadvantage of significantly disturbing the normal operation of the system. A method which makes use of a stationary random input signal and correlation functions is nondisturbing but the time required for identification is, in general, rather protracted.

In the present paper, a method for linear system identification is proposed which is based on the error-correcting training procedure in learning machines,[1,2] and is referred to as "learning identification."

II. Principle of Learning Identification

Considerations are restricted to systems with discrete time (sampled-data system) for the convenience of computer application, and noise-free measurements are assumed throughout.

A sampled weighting function of the linear time-invariant stable system is approximated by a finite set of values

$$w_1, w_2, w_3, \cdots, w_N, \tag{1}$$

where w_i is the ith sampled impulse response of the system, and N is so chosen that $N\Delta t$ covers the significant duration of impulse response, Δt is the sampling period determined by the sampling theorem. Let an input sequence be

$$x_1, x_2, x_3, \cdots, \tag{2}$$

and the corresponding output sequence of a system specified by (1) be

$$y_1, y_2, y_3, \cdots. \tag{3}$$

Then, for $j \geq N+1$, it follows that

$$y_j = w_1 x_{j-1} + w_2 x_{j-2} + \cdots + w_N x_{j-N} = \sum_{i=1}^{N} w_i x_{j-i}. \tag{4}$$

Thus, the problem of identifying a system specified by (1) is to find the sequence of a set of values

$$v_1^{(j)}, v_2^{(j)}, v_3^{(j)}, \cdots, v_N^{(j)} \quad (j = N+1, N+2, \cdots) \tag{5}$$

by an iterative procedure, such that each $v_i^{(j)}$ approaches w_i with the iteration step j. Such a set of values is called the weighting function of the identifier, and the output z_j of the identifier approaches y_j with the iteration step j, where

$$\begin{aligned} z_j &= v_1^{(j)} x_{j-1} + v_2^{(j)} x_{j-2} + \cdots + v_N^{(j)} x_{j-N} \\ &= \sum_{i=1}^{N} v_i^{(j)} x_{j-i}. \end{aligned} \tag{6}$$

Define the following N-dimensional vectors:

$$\mathbf{w} \triangleq \begin{pmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_N \end{pmatrix}, \qquad \mathbf{v}_j \triangleq \begin{pmatrix} v_1^{(j)} \\ v_2^{(j)} \\ \cdot \\ \cdot \\ \cdot \\ v_N^{(j)} \end{pmatrix}, \qquad \mathbf{x}_j \triangleq \begin{pmatrix} x_{j-1} \\ x_{j-2} \\ \cdot \\ \cdot \\ \cdot \\ x_{j-N} \end{pmatrix}.$$

The last vector is a modified input obtained by collecting N terms of the original input sequence (2). Hereafter, the vector input sequence

$$\mathbf{x}_{N+1}, \mathbf{x}_{N+2}, \mathbf{x}_{N+3}, \cdots \tag{7}$$

is used instead of (2). By the use of these vectors,

$$y_j = (\mathbf{w}, \mathbf{x}_j), \tag{8}$$

$$z_j = (\mathbf{v}_j, \mathbf{x}_j). \tag{9}$$

Now the adjustment procedure for the weighting function $\mathbf{v}_j$ of the identifier is as follows. The identification error for an input vector is allotted to each component of the weighting function vector of the identifier, proportional to the magnitude of the corresponding component of the input vector so that the output of the adjusted identifier gives a correct output if the same input is applied at the next sampling instant. More

Manuscript received April 11, 1966; revised September 26, 1966 and January 25, 1967. This work was supported by the Toyo Rayon Science Foundation.

The authors are with the Dept. of Mathematical Engrg. and Instrumentation Physics, University of Tokyo, Tokyo, Japan.

[1] B. Widrow, "Adaptive sampled-data systems," *1959 IRE WESCON Conv. Rec.*, pt. 4, pp. 74–85.

[2] H. M. Martinez, "A convergence theorem for linear threshold elements," *Bull. Math. Biophys.*, vol. 27, pp. 153–159, June 1965.

precisely, at the jth step, Δv_j is added, to v_j, where

$$\Delta v_j = v_{j+1} - v_j = (y_j - z_j)\frac{x_j}{\|x_j\|^2},$$

$$\left(\|x_j\|^2 \triangleq \sum_{i=1}^{N} x_{j-i}^2\right) \quad (10)$$

which means an error-correcting procedure in the sense that

$$(v_{j+1}, x_j) = (v_j + \Delta v_j, x_j) = (v_j, x_j) + (\Delta v_j, x_j)$$
$$= z_j + (y_j - z_j) = y_j. \quad (11)$$

The next task, then, is to show that the sequence $\{v_j\}$ converges to w independent of the initial value v_{N+1} for a random input sequence $\{x_j\}$ or $\{x_j\}$.

A geometrical interpretation of this adjustment procedure is as follows (see Fig. 1). Define a hyperplane Π_{j-1} in the N-dimensional space by

$$\Pi_{j-1} \triangleq \{p \mid (p, x_{j-1}) = y_{j-1}\}. \quad (12)$$

The hyperplane Π_{j-1} is perpendicular to vector x_{j-1} and, from (8) and (11), it is known that points w and v_j are on Π_{j-1}. Similarly, hyperplane Π_j is perpendicular to x_j, and points w and v_{j+1} are on Π_j. Since, from (10), vector $(v_{j+1}-v_j)$ is parallel to vector x_j, the point v_{j+1} coincides with the foot of the normal from point v_j to hyperplane Π_j.

Hence, letting the angle between the two vectors x_j and $u_j \triangleq w - v_j$ be θ_j,

$$\frac{\|u_{j+1}\|^2}{\|u_j\|^2} = \sin^2\theta_j = 1 - \cos^2\theta_j, \quad (13)$$

where

$$\cos^2\theta_j = \left(\frac{(x_j, u_j)}{\|x_j\|\|u_j\|}\right)^2, \qquad \|u_j\|^2 \triangleq \sum_{i=1}^{N} u_{j-i}^2. \quad (14)$$

Thus

$$\|u_{j+1}\|^2 = \|u_{N+1}\|^2 \prod_{r=N+1}^{j} (1 - \cos^2\theta_r), \quad (15)$$

and

$$\|u_j\| \to 0 \qquad (j \to \infty) \quad (16)$$

for an appropriate input sequence $\{x_j\}$. A condition for the convergence is given in Appendix I.

The adjustment procedure mentioned above can be slightly extended. Let

$$\Delta v_j = v_{j+1} - v_j = \alpha(y_j - z_j)\frac{x_j}{\|x_j\|^2} \quad (17)$$

instead of (10), where α is an error-correcting coefficient, and $2 > \alpha > 0$.

The learning method for system identification which makes use of (17) as the adjustment procedure is named the "fundamental method."

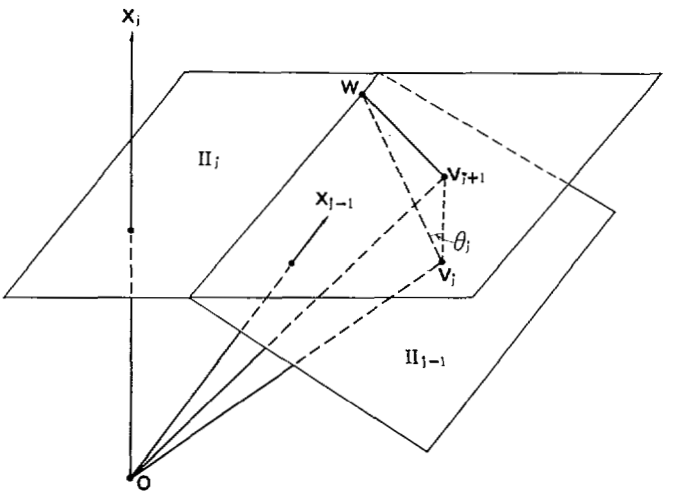

Fig. 1. Geometrical interpretation of the adjustment procedure when $N=3$.

III. Computer Simulation of Learning Identification by the Fundamental Method

Suppose that input sequence (2) is applied to a discrete linear system represented by (1). Hereafter, the initial value of the weighting function vector of the identifier is assumed to be a zero vector.

The normalized error e_j of the identification at the jth step is defined by

$$e_j \triangleq \frac{\|w - v_j\|}{\|w\|} \qquad (j \geqq N+1). \quad (18)$$

It is obvious from (13) that e_j is a nonincreasing function of j. An integer $j(\varepsilon)$ is denoted as the first number j in such a manner that the error e_j becomes equal to or less than a given positive number ε; $j(\varepsilon)$ is a nonincreasing function of ε. Normalized by the extent of the weighting function N, $T(\varepsilon)$ is defined as

$$T(\varepsilon) \triangleq \frac{j(\varepsilon)}{N} - 1. \quad (19)$$

The reason for subtracting 1 is that $e_j = 1$ at $j = N+1$. Note that $T(\varepsilon)$ also is a nonincreasing function of ε, and is called "time for identification" or "identification time."

A set of values $w_1, w_2, \cdots, w_N$ representing the sampled weighting function of the unknown system was set in the computer. Even though computer simulations were carried out for various systems with different weighting functions, the results concerning a sampled weighting function which corresponds to the second-order system:

$$\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

with $\zeta = 0.5$ and $\omega_n = 1$ will be described in the following, because no significant difference was found in the variety of weighting functions (see Table III). The sampling period Δt may be assumed to be arbitrary, because

the time scale was normalized by Δt in the computer simulation.

As for the input, stationary random sampled signals with Gaussian distribution, which were generated by computational programming, were used unless otherwise stated, although it is not necessary for the signals to be stationary or Gaussian.

An example of the process of convergence is illustrated in Fig. 2.

In the case where $N=16$, the influence of α on $T(\varepsilon)$ is shown in Table I. From this, the identification time is minimum in the neighborhood of $\alpha=1$. (Compare with the case of the "quantizing method" appearing later.)

Table II shows the relation N vs. $T(\varepsilon)$, when $\alpha=1$. $T(\varepsilon)$ may be regarded as independent of N, provided $N\geq 8$. In the case of $\alpha=1$ and $N=16$, results of the identifications of a second-order system with $\zeta=0.5$, $\omega_n=1$ and of a first-order system with a dead time of 3 sampling intervals are compared in Table III.

A comparison is made in Table IV where random signals with Gaussian distribution and uniform distribution were used for the input. No significant difference is observable between these two cases.

Any given system can be identified by the same procedure as above, using a sequence of independent random vectors as the input sequence, even though, in the preceding discussion, the input vector $\mathbf{x}_j$ was dependent on $\mathbf{x}_{j-1}$, $\mathbf{x}_{j-2}$, $\cdots$, $\mathbf{x}_{j-N+1}$ (see Appendix II). Results of simulation in the above independent case are shown in Table V in comparison with those of the dependent case. In both cases, N components of each input vector are independent random variables with the same nominal distribution. From those results, it seems likely that the convergence of identification is slightly faster in the dependent case than in the independent case (apart from factor N). In the independent case, the expected value of the identification time can be easily calculated from Appendix II (see third row of Table V). As shown,

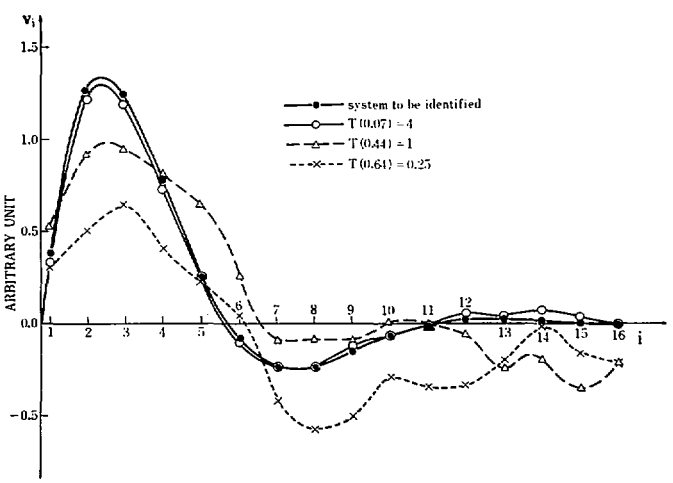

Fig. 2. Identification by the fundamental method ($\alpha=1$) of a second-order system ($\zeta=0.5$, $\omega_n=1$; $N=16$), the input of which is a random signal with normal distribution, was simulated on a digital computer. Several weighting functions of the identifier in the course of convergence are shown.

TABLE I

IDENTIFICATION BY THE FUNDAMENTAL METHOD OF A SECOND-ORDER SYSTEM FOR VARIOUS VALUES OF α

α \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
0.01	more than 63		
0.05	more than 63		
0.2	15.5	24	63
0.5	5.4	8.2	22.1
0.75	4.5	6.5	16.8
1.0	4.2	6.1	13.6
1.25	3.6	6.0	19.6
1.5	5.6	9.0	26.9
1.75	14.5	22.1	53

TABLE II

IDENTIFICATION BY THE FUNDAMENTAL METHOD ($\alpha=1$) OF A SECOND-ORDER SYSTEM FOR VARIOUS VALUES OF N

N \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
4	1.7	2.4	9.2
8	2.9	5.0	15.9
16	4.2	6.1	13.6
32	3.5	5.4	15.5
64	4.0	6.5	15.6

TABLE III

IDENTIFICATION BY THE FUNDAMENTAL METHOD ($\alpha=1$)

Systems \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
1st-Order System with Dead Time	4.1	6.4	13.8
2nd-Order System with $\zeta=0.5$, $\omega_n=1$	4.2	6.1	13.6

TABLE IV

IDENTIFICATION BY THE FUNDAMENTAL METHOD ($\alpha=1$) OF A SECOND-ORDER SYSTEM

Random Input Signals \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
Normal Distribution	5.1	7.0	16.7
Uniform Distribution	4.4	7.3	17.3

TABLE V

IDENTIFICATION BY THE FUNDAMENTAL METHOD ($\alpha=1$) OF A SECOND-ORDER SYSTEM

$N=16$	$T(\varepsilon)$
Random Input Vectors \ ε	5%	1%	10^{-5}
Dependent (Experimental)	5.1	7.0	16.7
Independent (Experimental)	$5.3\times N$	$8.0\times N$	$20.5\times N$
Independent (Theoretical)	$6.0\times N$	$9.2\times N$	$23.0\times N$

the expected times for various values of ε coincide fairly well with the results of the computer simulation.

IV. Some Modifications of the Fundamental Method of Learning Identification

A. Quantizing Method

With the input signals quantized into two levels, the following adjustment procedure is adopted

$$\Delta v_j = v_{j+1} - v_j = \beta(y_j - z_j)\frac{\mathbf{sgn}\, x_j}{|x_j|}, \qquad (\beta > 0) \qquad (20)$$

instead of (17), where $\mathbf{sgn}\, x_j$ is a vector whose components are the signs of the corresponding components of the vector x_j, and

$$|x_j| \triangleq \sum_{i=1}^{N} |x_{j-i}| = (\mathbf{sgn}\, x_j, x_j). \qquad (21)$$

This method is called the "quantizing method."

Identification by the quantizing method of a second-order system ($\zeta = 0.5$, $\omega_n = 1$; $N = 16$) was simulated using a digital computer. The influence of the variations of the error-correcting coefficient β on $T(\varepsilon)$ is shown in Table VI. Unlike the case of the fundamental method, the identification time was minimum when $\beta = 0.5 \sim 0.75$, and the process of identification did not converge if $\beta \geq 1.5$.

Table VII shows the relation N vs. $T(\varepsilon)$, when $\zeta = 0.5$, $\omega_n = 1$, and $\beta = 1$.

Although the quantizing method requires a longer identification time than the fundamental method, it has the advantage that the hardware realization becomes much simpler.

TABLE VI
Identification by the Quantizing Method of a Second-Order System for Various Values of β

β \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
0.05	58.1	more than 63	
0.2	14.4	22.1	63.0
0.5	6.3	10.9	29.7
0.75	5.3	12.0	30.1
1.0	8.5	17.9	37.1
1.25	44.3	more than 63	
1.5	∞		
1.75	∞		

TABLE VII
Identification by the Quantizing Method ($\beta = 1$) of a Second-Order System for Various Values of N

N \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
4	4.5	12.0	18.2
8	8.7	17.5	45.5
16	8.5	17.9	37.1
32	12.6	17.4	more than 31.3
64	7.9	14.6	more than 15.6

B. Repeating Method

In order to reduce the identification time, an effective method is to perform the adjustment procedure repeatedly. This is called the "repeating method." As the input, $m-1$ input vectors

$$x_{j-1}, x_{j-2}, \cdots, x_{j-m+1}$$

are held during the jth step beside the input vector x_j of the fundamental method. Moreover, $m-1$ outputs

$$y_{j-1}, y_{j-2}, \cdots, y_{j-m+1}$$

are held beside the output y_j of the fundamental method. Unlike a once-for-all operation of the adjustment procedure at each step in the fundamental method, in this method a series of m differing adjustment procedures

$$\Delta v_j^{(k,h)} = (y_{j-k+1} - z_{j-k+1}^{(h)})\frac{x_{j-k+1}}{\|x_{j-k+1}\|^2}$$
$$(k = 1, 2, \cdots, m;\ h = 1, 2, \cdots, l) \qquad (22)$$

is repeated at the jth step.

Identification by the repeating method of a second-order system ($\zeta = 0.5$, $\omega_n = 1$; $N = 16$), the input of which is a Gaussian random signal, was simulated on a digital computer. The results are shown in Table VIII.

The disadvantage of this method is that a slightly larger memory content must be provided in order to hold the extra input and output signals, and that each adjustment procedure has to be completed within a short time compared to the sampling interval; while its great advantage is that the identification time is considerably reduced by making use of the time before the next sampling. The identification has been improved even if the repetition is stopped halfway.

TABLE VIII
Identification by the Repeating Method of a Second-Order System for Various Values of l and m

l	m \ ε	$T(\varepsilon)$ 5%	1%	10^{-5}
2	4	3.7	4.9	12.5
	8	2.6	4.0	8.1
	16	0.88	1.44	2.5
4	16	0.48	0.88	1.94
8	4	3.6	4.8	12.1
	8	2.1	3.1	7.3
	16	0.37	0.5	1.56
	32	0.06	0.13	0.44
16	16	0.27	0.34	0.87
32	16	0.13	0.25	0.69
64	4	3.7	4.8	12.1
	8	2.1	3.1	7.3
	16	0.06	0.13	0.5

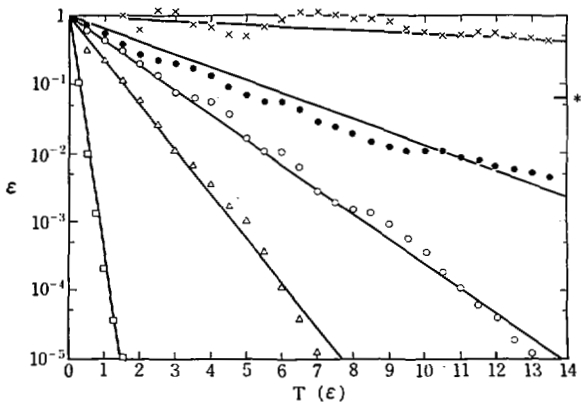

Fig. 3. Some typical examples of the time for identification by various methods of the learning identification using a random signal with normal distribution.

—○— by the fundamental method ($\alpha=1$)
—●— by the quantizing method ($\beta=0.5$)
—△— by the repeating method of the fundamental procedure ($m=8$, $l=8$)
—□— by the repeating method of the fundamental procedure ($m=16$, $l=8$)
—×— by the method using correlation functions (* indicates the error at $T=100$).

Some typical examples of the identification are shown in Fig. 3, as well as the method using correlation functions with no noise present. Although a second-order system with $\zeta=0.5$, $\omega_n=1$, and $N=16$ was chosen as the system to be identified, the results were not peculiar to the second-order system.

V. Conclusion

A learning method for linear time-invariant system identification, which uses random signals (not necessarily stationary) as the input, was proposed. The method is based on an iterative procedure which is similar to the error-correcting training procedure in learning machines, and it was shown theoretically and experimentally that the iteration converges for a wide class of input signals.

The identification can be completed within a much shorter time by this method than by correlation functions, and will be applicable with a certain accuracy to the identification of quasi-time-invariant systems, in which some parameters vary slowly in comparison with the time required for identification.

The method was simulated by a digital computer to investigate the above-mentioned merits, and the data for hardware realization were obtained for various cases. Some modifications of the method were also presented, together with their computer simulations.

Appendix I
Convergence of Learning Identification

Equation (17) is rewritten in the form

$$v_{j+1} = v_j + \alpha(w - v_j, x_j)\frac{x_j}{\|x_j\|^2} = v_j + \alpha\frac{x_j x_j'}{\|x_j\|^2}(w - v_j). \qquad (2 > \alpha > 0) \qquad (23)$$

Since $u_j = w - v_j$, (23) becomes

$$u_{j+1} = X_j u_j \qquad (j \geq N+1), \qquad (24)$$

and

$$X_j = I - \alpha\frac{x_j x_j'}{\|x_j\|^2}, \qquad (25)$$

where I is the $N\times N$ unit matrix. Thus the problem is reduced to show that $u_j \to 0$ or $\|u_j\| \to 0$ when $j \to \infty$.

From (24), it follows

$$\|u_{j+1}\|^2 = (X_j u_j, X_j u_j) = (u_j, X_j^2 u_j) = (1 - \xi_j)\|u_j\|^2, \qquad (26)$$

where

$$\xi_j = \alpha(2 - \alpha)\left(\frac{(x_j, u_j)}{\|x_j\|\|u_j\|}\right)^2, \qquad (27)$$

and

$$\|u_{j+1}\|^2 = \|u_{N+1}\|^2 \prod_{r=N+1}^{j}(1 - \xi_r). \qquad (28)$$

Since $1 \geq \xi_r \geq 0$ for $2 > \alpha > 0$, necessary and sufficient condition for $\|u_j\| \to 0$ as $j \to \infty$, is given by

$$\sum_{j=N+1}^{\infty} \xi_j = \infty. \qquad (29)$$

Now if the input sequence (2) is a random sequence, the input vector x_j is a random (vector) variable, and hence

$$\xi_{N+1}, \xi_{N+2}, \xi_{N+3}, \cdots \qquad (30)$$

is a sequence of random variables, where ξ_j is dependent on

$$\xi_{N+1}, \xi_{N+2}, \xi_{N+3}, \cdots, \xi_{j-1}.$$

Put

$$\eta_j \triangleq E\{\xi_j \mid \xi_{N+1}, \xi_{N+2}, \cdots, \xi_{j-1}\}, \qquad (31)$$

and assume that

$$\sum_{j=N+1}^{\infty} \eta_j = +\infty. \qquad (32)$$

Then (29) follows from Borel-Cantelli's extended lemma,[3,4] and it may be concluded that (32) is necessary and sufficient for the convergence:

$$\|u_j\| \to 0 \qquad (j \to \infty). \qquad (33)$$

For example, the assumption that in sequence (30) there occurs an infinity of ξ_j's such that

$$\eta_j > \delta,$$

where δ is a positive constant, is sufficient for (33).

[3] This was suggested by S. Sato.
[4] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1963, p. 323 (Corollary 1).

Appendix II
Estimation of Identification Time

Let x_k's in the input sequence (2) be independent random variables having the same symmetric distribution with respect to zero, and define the input vectors $x_j (j \geq 1)$ by

$$x_j \triangleq \text{col.}\,(x_{jN}, x_{jN-1}, x_{jN-2}, \cdots, x_{jN-N+1}). \tag{34}$$

Note that in this case, x_j's are independent. Then, for $j \geq 1$,

$$y_j \triangleq (w, x_j),$$
$$z_j \triangleq (v_j, x_j).$$

From (24),

$$\|u_{j+1}\|^2 = (X_j u_j, X_j u_j) = (u_j, X_j^2 u_j)$$
$$= u_j' X_j^2 u_j, \tag{35}$$

where

$$X_j^2 = I - \alpha(2-\alpha)\frac{x_j x_j'}{\|x_j\|^2}. \tag{36}$$

The last term at the right side of (36) is a matrix whose (p, q) element is

$$r_{pq} = \frac{x_{jN-p+1} x_{jN-q+1}}{\rho} \qquad (p, q = 1, 2, \cdots, N), \tag{37}$$

where

$$\rho = x_{jN}^2 + x_{jN-1}^2 + \cdots + x_{jN-N+1}^2. \tag{38}$$

Since x_{jN-p+1}'s $(p = 1, 2, \cdots, N)$ are assumed to be independent, identically distributed, and symmetrically distributed, it is easy to see that

$$\bar{r}_{pq} = \begin{cases} 0 & (p \neq q), \\ N^{-1} & (p = q). \end{cases} \tag{39}$$

Consequently, from (36),

$$\overline{X^2} = \{1 - \alpha(2-\alpha)N^{-1}\} I. \tag{40}$$

Note that the random vectors u_j and x_j are independent, so that u_j and X_j^2 are also independent, and it follows from (35) and (40) that

$$\overline{\|u_{j+1}\|^2} = \{1 - \alpha(2-\alpha)N^{-1}\}\overline{\|u_j\|^2}, \tag{41}$$

or

$$\overline{\|u_{j+1}\|^2} = \{1 - \alpha(2-\alpha)N^{-1}\}^j \|u_1\|^2. \tag{42}$$

Letting $v_1 = 0$ as before, define the "expected normalized error" $\bar{e}_j$ at the jth step by the rms value of the normalized error; that is, by

$$\bar{e}_j \triangleq \sqrt{\overline{\left(\frac{\|w - v_j\|}{\|w\|}\right)^2}}. \tag{43}$$

For a given positive number ε, denote the first number j such that $\bar{e}_j$ becomes equal to or less than ε as $\overline{T}(\varepsilon)$ and call it "expected time for identification." (Note that this has already been normalized by N, as seen from the definition of x_j.)

Calculating $\overline{T}(\varepsilon)$ from (42),

$$\overline{T}(\varepsilon) \doteqdot \frac{2N \ln \varepsilon^{-1}}{\ln\{1 - \alpha(2-\alpha)N^{-1}\}^{-N}} + 1. \tag{44}$$

When N is large and ε is small (for example, $N \geq 10$, $\varepsilon \leq 0.1$), (44) is approximated by

$$\overline{T}(\varepsilon) \doteqdot \frac{2N}{\alpha(2-\alpha)} \ln \varepsilon^{-1}. \tag{45}$$

Some numerical data obtained by using (45) are shown in Table V in comparison with the experimental data. The coincidence between them is fairly satisfactory. Also, data in Table I seem to confirm (45) aside from factor N. It is worth mentioning that (45) holds, regardless of the form of the weighting function to be identified.

Finally, consider the case where the quantizing method with $2 > \beta > 0$ is used. In this case

$$X_j^2 = I - \beta(2-\beta)\frac{(\operatorname{sgn} x_j) x_j'}{|x_j|} \tag{46}$$

is obtained instead of (36), and

$$\overline{T}(\varepsilon) \doteqdot \frac{2N}{\beta(2-\beta)} \ln \varepsilon^{-1}. \tag{47}$$

This explains the data in Table VI qualitatively but not quantitatively.