XIX. Communications Systems Research: Propagation Studies

TELECOMMUNICATIONS DIVISION

A. Two Stochastic Approximation Procedures for Identifying Linear Systems, *J. K. Holmes*

1. Introduction

Many important problems in communications research can be posed as a problem of system identification. That is, given an input signal record and an output signal record, find an equivalent system that fits these data. An important example occurs in the following communication problem where it is desired to study the effects of the atmosphere on signal transmission. A transmitter transmits directly to a tracking station via cables or a microwave link, and to a spacecraft which retransmits to the tracking station. The signal, as it passes through, for example, the Solar Corona to and from the spacecraft, undergoes distortion. One method of studying the distortion is to characterize it as a finite memory linear system. The identification could be derived from the input (the directly transmitted component) with the proper time delay, and the output, which is the retransmitted, distorted signal. Naturally, noise would, in general, corrupt both the input and output signal measurements.

This article, then, considers the basic problem of identifying linear systems from noisy input–output measurements. Related work is listed in Refs. 1–5.

Let N denote the set of natural numbers and let n be in N. Denote by $\mathcal{S}_L(\ell)$ the class of linear, finite memory, time invariant, time discrete systems. Then, for systems $s \in \mathcal{S}_L(\ell)$, we can relate the input and output random sequences by

$$v_n = \sum_{j=0}^{\ell-1} \phi_j u_{n-j}, \qquad \ell < \infty \tag{1}$$

The vector $\boldsymbol{\phi} = (\phi_0, \cdots, \phi_{\ell-1})$ defines the system when it is in the class $\mathcal{S}_L(\ell)$. Further, we assume that the observable sequences v'_n and u'_n are defined as $v'_n = v_n + \delta_n$ and $u'_n = u_n + \epsilon_n$, where ϵ_n and δ_n are mutually and individually independent noise sequences with zero means and respective variances σ_ϵ^2 and σ_δ^2. Also, u_n is assumed to be independent of ϵ_n and δ_n. This article, then, is concerned with the following problem: From the noise-corrupted input and output measurements, estimate the unknown system (i.e., $\boldsymbol{\phi}$) via nonparametric methods. See Fig. 1 for a block diagram.

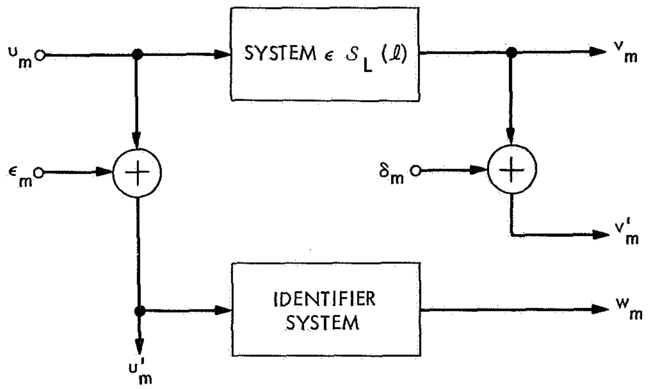

Fig. 1. Block diagram of available measurements u'_m and v'_m and conceptual location of identifier system

2. Development of the First Algorithm

Let the estimator of the sequence $v_m(\boldsymbol{\phi})$ be

$$w_m(\mathbf{z}) = \sum_{j=0}^{\ell-1} z_j u'_{m-j}$$

where $m = \ell n$ and the row vector $\mathbf{z} \in E^\ell$. Now, we shall derive a specific sequential estimate of $\boldsymbol{\phi}$ denoted by $\mathbf{x}^m$. Define an error measure in the following way:

$$M(\mathbf{z}) = E\left[(w_m(\mathbf{z}) - v'_m)^2 \,|\, \mathbf{z}\right] \tag{2}$$

The procedure to be used to estimate $\boldsymbol{\phi}$ will be to recursively determine $\mathbf{z}$ in such a manner that $M(\mathbf{z})$ is a minimum. Specifically, we consider a Kiefer–Wolfowitz stochastic approximation method.

Let

$$\rho_m = (w_m(\mathbf{x}^m) - v'_m)^2$$

and define

$$D_i^m = \rho_m(\mathbf{x}^m + c_m \mathbf{e}_i) - \rho_m(\mathbf{x}^m - c_m \mathbf{e}_i), \quad i = 0, \cdots, \ell - 1$$

where c_m satisfies

$$c_m \geqq 0 \qquad \text{and} \qquad \lim_{m\to\infty} c_m = 0$$

and the $\mathbf{e}_i$ are the orthonormal unit vectors $\in E^\ell$, viz.,

$$\mathbf{e}_0 = (1, 0, \cdots, 0) \qquad \text{and} \qquad \mathbf{e}_{\ell-1} = (0, 0, \cdots, 1)$$

Then, recursively define $\mathbf{x}^m$ by

$$\mathbf{x}^{m+\ell} = \mathbf{x}^m - a_m \frac{\mathbf{D}^m}{c_m} \text{for } m = \ell, 2\ell, 3\ell, \cdots \tag{3}$$

where

$$\mathbf{D}^m = (D_0^m, D_1^m, \cdots, D_{\ell-1}^m)$$

and a_m satisfies

$$a_m \geqq 0 \qquad \text{and} \qquad \lim_{m\to\infty} a_m = 0$$

By using the definition of $\mathbf{D}^m$ and ρ_m, one obtains the following scalar algorithm for each component of the estimate:

$$x_i^{m+\ell} = x_i^m - 4a_m u'_{m-i}\left[\sum_{j=0}^{\ell-1} x_j^m u'_{m-j} - v'_m\right], \quad i = 0, \cdots, \ell - 1$$

or in vector form, designating $(\mathbf{u}'_m)^T$ as the transpose of $\mathbf{u}'_m$, we have

$$\mathbf{x}^{m+\ell} = \mathbf{x}^m - 4a_m \mathbf{x}^m (\mathbf{u}'_m)^T \mathbf{u}'_m + 4a_m v'_m \mathbf{u}'_m \tag{4}$$

which is independent of c_m, and where

$$\mathbf{u}'_m = (u'_m, u'_{m-1}, \cdots, u'_{m-\ell+1})$$

3. The Estimate Error

Based on minimizing the mean square error between the sequence $v_m(\boldsymbol{\phi})$ and $w_m(\mathbf{x}^m)$, where $\mathbf{x}^m$ is defined by Eq. (3), we show that $\mathbf{x}^m$ does not converge to $\boldsymbol{\phi}$ unless the input noise sequence ϵ_n is identically zero. More precisely, we have our first result.

Theorem 1. If the conditions

(a) $a_m \geqq 0$, $\sum_1^\infty a_m = \infty$, and $\sum_1^\infty a_m^2 < \infty$.

(b) $\mathbf{y}R(\mathbf{y})^T > 0$, $\forall \mathbf{y} \neq 0$, and $E[\mathbf{u}_m^T(\mathbf{u}_m) | \mathbf{x}^m] = \mathrm{R}$.

(c) $E[(u'_m)^4 | \mathbf{x}^m] < c_1 < \infty$, and $E[u_m^2 \delta_m^2 | \mathbf{x}^m] < c_2 < \infty$.

(d) The random sequence u_m and both noise sequences δ_m and ϵ_m are time-stationary.

(e) $E[\delta_m] = E[\epsilon_m] = 0, E[\delta_m \delta_n] = \sigma_\delta^2 \delta_{mn}$, and $E[\epsilon_m \epsilon_n] = \sigma_\epsilon^2 \delta_{mn}$.

(f) $E[\delta_m \epsilon_n] = E[u_m \delta_n] = E[\delta_m \epsilon_n] = 0$.

(g) Unknown system $s \epsilon \mathcal{S}_L(\ell)$.

are met, then the sequence $\mathbf{x}^m$, defined by Eq. (3), converges in mean square to the vector

$$\boldsymbol{\theta} = [R + \sigma_\epsilon^2 I]^{-1} R \boldsymbol{\phi}$$

Proof. Equation (4) may be expanded to

$$\mathbf{x}^{m+1} = \mathbf{x}^m - 4a_m \mathbf{u}'_m (\mathbf{u}'_m)^T \mathbf{x}^m + 4a_m [\delta_m \mathbf{u}'_m + \mathbf{u}'_m (\mathbf{u}_m)^T \boldsymbol{\phi}] \qquad (5)$$

where in Eq. (5) and throughout the rest of this proof $\mathbf{x}^m$, $\mathbf{u}'_m$, $\mathbf{u}_m$, and $\boldsymbol{\phi}$ are now defined as column vectors instead of row vectors. Using $E[(\cdot)] = EE[(\cdot)|\mathbf{x}^n]$ on both sides of Eq. (5) results in

$$E[\mathbf{x}^{m+1}] = E[\mathbf{x}^m] - 4a_m [R + \sigma_\epsilon^2 I] E[\mathbf{x}^m] + 4a_m R\boldsymbol{\phi} \qquad (6)$$

Clearly, a solution to Eq. (6) is given by

$$E[\mathbf{x}^m] = [R + \sigma_\epsilon^2 I]^{-1} R\boldsymbol{\phi} \qquad (7)$$

We shall now show that

$$\lim_{m \to \infty} E[\|\mathbf{x}^m - \boldsymbol{\theta}\|^2] = 0$$

Subtracting $\boldsymbol{\theta}$ from both sides of Eq. (5) and rearranging yields

$$\begin{aligned}(\mathbf{x}^{m+1} - \boldsymbol{\theta}) = (\mathbf{x}^m - \boldsymbol{\theta}) &- 4a_m \mathbf{u}'_m (\mathbf{u}'_m)^T (\mathbf{x}^m - \boldsymbol{\theta}) \\ &+ 4a_m [\delta_m \mathbf{u}'_m + \mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta}]\end{aligned}$$

Forming the norm squares of both sides and averaging, we obtain

$$\begin{aligned} b_{m+1} = b_m &- 8a_m E \langle \mathbf{x}^m - \boldsymbol{\theta}, \mathbf{u}'_m (\mathbf{u}'_m)^T (\mathbf{x}^m - \boldsymbol{\theta}) \rangle \\ &+ 16a_m^2 E[\|\mathbf{u}'_m (\mathbf{u}'_m)^T (\mathbf{x}^m - \boldsymbol{\theta})\|^2] \\ &+ 16a_m^2 E[\|\delta_m \mathbf{u}'_m + \mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta}\|^2] \\ &+ 8a_m E \langle \mathbf{x}^m - \boldsymbol{\theta}, \delta_m \mathbf{u}'_m + \mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta} \rangle \\ &- 32a_m^2 E \langle \mathbf{u}'_m (\mathbf{u}'_m)^T (\mathbf{x}^m - \boldsymbol{\theta}), \delta_m \mathbf{u}'_m \\ &+ \mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta} \rangle \end{aligned} \qquad (8)$$

where we have let

$$b_m = E[\|\mathbf{x}^m - \boldsymbol{\theta}\|^2]$$

By Condition (b) we have for the second term in Eq. (8)

$$\sigma_\epsilon^2 b_m < \lambda_1 b_m \leqq \langle \mathbf{x}^m - \boldsymbol{\theta}, [R + \sigma_\epsilon^2 I] (\mathbf{x}^m - \boldsymbol{\theta}) \rangle \qquad (9)$$

where λ_1 is the minimum eigenvalue of $[R]$. The third term can be bounded, using the Schwartz inequality, by

$$16a_m^2 b_m k_1 \qquad (10)$$

where $k_1 < \infty$ by Condition (c). The fourth term can be bounded by the following quantity

$$32a_m^2 E[\|\delta_m \mathbf{u}'_m\|^2] + 32a_m^2 E[\|\mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi}\|^2] + 32a_m^2 E[\|\mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta}\|^2]$$

All these terms are similarly bounded by

$$k_2 a_m^2 \qquad (11)$$

where $k_2 < \infty$. Now, the fifth term can be reduced to

$$8a_m E \langle \mathbf{x}^m - \boldsymbol{\theta}, R\boldsymbol{\phi} - [R + \sigma_\epsilon^2 I] \boldsymbol{\theta} \rangle = 0$$

since $\boldsymbol{\theta} = [R + \sigma_\epsilon^2 I]^{-1} R\boldsymbol{\phi}$. The last term is bounded by

$$\begin{aligned} &32a_m^2 \{E[\|\mathbf{u}'_m (\mathbf{u}'_m)^T (\mathbf{x}^m - \boldsymbol{\theta})\|^2] E[\|\mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_m (\mathbf{u}'_m)^T \boldsymbol{\theta}\|^2]\}^{1/2} \\ &\leqq 32a_m^2 \{E[\|\mathbf{x}^m - \boldsymbol{\theta}\|^2 E[\|\mathbf{u}'_m (\mathbf{u}'_m)^T\|^2 | \mathbf{x}^m]] E[\|\mathbf{u}'_m \mathbf{u}_m^T \boldsymbol{\phi} - \mathbf{u}'_\mathrm{m} (\mathbf{u}'_m)^T \boldsymbol{\theta}\|^2]\}^{1/2} \end{aligned} \qquad (12)$$

where $\|A\|$, with A a square matrix, is the usual euclidean ℓ_2 norm; i.e.,

$$\|A\| = (\sum_{i,j} a_{ij}^2)^{1/2}$$

Equation (12) can be shown to be bounded by

$$32a_m^2 (b_m k_3)^{1/2} \leqq a_m^2 (1 + b_m) k_4 \qquad (13)$$

for $k_4 < \infty$. Hence, Eq. (8) yields

$$\begin{aligned} b_{m+\ell} &\leqq b_m \left[1 - 8a_m (\lambda_1 - k_5 a_m)\right] + a_m^2 k_6 \\ &\leqq b_m (1 - 4a_m \lambda_1) + a_m^2 k_6 \qquad (14) \end{aligned}$$

for m sufficiently large and k_5 and k_6 finite. By applying an interesting application of Kronecker's theorem (e.g., Ref. 5) or Lemma I of Ref. (3), we have that $b_m \rightarrow 0$. The theorem follows directly.

To conclude the statement made in the first paragraph of this subsection, we have:

Lemma 1. If the conditions of Theorem 1 are satisfied and if $\boldsymbol{\phi} \neq 0$, then $\mathbf{x}^m$ is a consistent estimator of $\boldsymbol{\phi}$ if and only if $\sigma_\epsilon^2 = 0$.

Proof. The Lemma follows from Theorem 1 and the fact that $[R + \sigma_\epsilon^2 I]^{-1} R\boldsymbol{\phi}$ is equal to $\boldsymbol{\phi}$ if and only if $\sigma_\epsilon^2 = 0$, since by Condition (b) R is positive definite.

The fact that $\mathbf{x}^m$ converges to $\boldsymbol{\theta} \neq \boldsymbol{\phi}$, based on minimizing $M(\mathbf{x}^m)$, is not so surprising since Eq. (2) can be written as

$$M(\mathbf{z}) = E\left[\left(\sum_{j=0}^{\ell-1} (z_j - \phi_j)\, u_{m-j}\right)^2 \Big| \mathbf{z}\right] + \sigma_\epsilon^2 \|\mathbf{z}\|^2 + \sigma_\delta^2 \qquad (15)$$

which reduces to

$$\begin{aligned} M(z) = {} & (\mathbf{z} - \boldsymbol{\theta})\,[R + \sigma_\epsilon^2 I]\,(\mathbf{z} - \boldsymbol{\theta})^T \\ & + (\boldsymbol{\theta} - \boldsymbol{\phi})\, R\, (\boldsymbol{\theta} - \boldsymbol{\phi})^T + \sigma_\epsilon^2 \|\boldsymbol{\theta}\|^2 + \sigma_\delta^2 \qquad (16) \end{aligned}$$

where

$$\boldsymbol{\theta} = [R + \sigma_\epsilon^2 I]^{-1} R\, \boldsymbol{\phi}$$

Since R is positive definite, we have that the minimum occurs at $\mathbf{z} = \boldsymbol{\theta}$. Hence, we *cannot* exactly identify an unknown system, when the input noise variance is nonzero, without additional knowledge of the input noise statistics.

4. Second Algorithm

A modification of the original algorithm, defining a new sequence $\mathbf{y}^m$, has the property that if σ_ϵ^2 is known then the modified algorithm leads to the result that $y_m \rightarrow \boldsymbol{\phi}$ in mean square.

Theorem 2. If the conditions of Theorem 1 are fulfilled and if $\mathbf{y}^m$ is defined recursively by

$$\mathbf{y}^{m+\ell} = \mathbf{y}^m - 4a_m\left[\mathbf{u}_m' (\mathbf{u}_m')^T - \sigma_\epsilon^2 I\right] \mathbf{y}^m + 4a_m \mathbf{u}_m' v_m' \qquad (17)$$

then $\mathbf{y}^m$ converges in mean square to $\boldsymbol{\phi}$.

Proof. The proof follows the lines of Theorem 1 and will not be indicated here.

5. Convergence to a Subspace

Up to this point, the condition that R be positive definite was required when $\sigma_\epsilon^2 = 0$ to prove convergence (in mean square) to a unique point. It is to be noted that if $\sigma_\epsilon^2 > 0$ then it is not necessary to assume that R is positive definite. However, if it turns out that R is of rank $r < \ell$, then it can be proven that $\mathbf{x}^n$ converges to a subspace of E^ℓ. Obviously, this is a very weak form of convergence. However, looking at it philosophically, we cannot hope to do any better since if $x^n \epsilon\, \eta$, the null space of R, $M(z)$ will be minimum and that is the best we can do using the mean square error criterion. This then leads us to the statement of Theorem 3.

Theorem 3. If the Conditions (a), (c)–(g), $0 <$ rank $R = r < \ell$, and $\sigma_\epsilon^2 = 0$ are satisfied, then $R(\mathbf{x}^n - \boldsymbol{\phi})$ converges in mean square to zero.

Proof. Starting with Eq. (5), setting ϵ_m equal to zero, subtracting $\boldsymbol{\phi}$ from both sides of the equation, and premultiplying by R yields

$$\begin{aligned} R(\mathbf{x}^{m+\ell} - \boldsymbol{\phi}) = {} & R(x^m - \boldsymbol{\phi}) - 4a_m R\mathbf{u}_m \mathbf{u}_m^T (\mathbf{x}^m - \boldsymbol{\phi}) \\ & + 4a_m \delta_m R\mathbf{u}_m \end{aligned}$$

Forming the norm square of both sides and averaging, we obtain

$$\begin{aligned} d_{m+\ell} = d_m & - 8a_m E\,\langle R(\mathbf{x}^m - \boldsymbol{\phi}), R\mathbf{u}_m \mathbf{u}_m^T (\mathbf{x}^m - \boldsymbol{\phi})\rangle \\ & + 16a_m^2 E\left[\|R\mathbf{u}_m \mathbf{u}_m^T (\mathbf{x}^m - \boldsymbol{\phi})\|^2\right] \\ & + 16a_m^2 E\left[\delta_m^2 \|R\mathbf{u}_m\|^2\right] \\ & + 8a_m E\,\langle R(\mathbf{x}^m - \boldsymbol{\phi}), \delta_m R\mathbf{u}_m\rangle \\ & - 32a_m^2 E\,\langle R\mathbf{u}_m \mathbf{u}_m^T (\mathbf{x}^m - \boldsymbol{\phi}), \delta_m R\mathbf{u}_m\rangle \qquad (18) \end{aligned}$$

where we have let

$$d_m = E\left[\|R(\mathbf{x}^m - \boldsymbol{\phi})\|^2\right]$$

Now, even though it is no longer true that

$$E\langle \mathbf{x}^m - \boldsymbol{\varphi}, R(\mathbf{x}^m - \boldsymbol{\varphi})\rangle \geqq \lambda_1 E[\|\mathbf{x}^n - \boldsymbol{\varphi}\|^2]$$

it is true that

$$E\langle R(\mathbf{x}^m - \boldsymbol{\varphi}), RR(\mathbf{x}^m - \boldsymbol{\varphi})\rangle \geqq \lambda_\rho E[\|R(\mathbf{x}^m - \boldsymbol{\varphi})\|^2] = \lambda_\rho d_m \qquad (19)$$

where λ_ρ is the least nonzero eigenvalue of R.

The third term can be bounded by

$$16a_m^2 d_m K_7 \qquad (20)$$

where $K_7 < \infty$ by Condition (c). The fourth term can be bounded by

$$16K_8 a_m^2 \qquad (21)$$

The fifth and sixth terms are zero since $E[\delta_m] = 0$.

Hence, we have that

$$d_{m+1} \leqq d_m[1 - 8a_m(\lambda_\rho - K_7 a_m)] + 16K_8 a_m^2$$

or

$$d_{m+1} \leqq d_m(1 - 4a_m \lambda_\rho) + 16K_8 a_m^2 \qquad (22)$$

for m sufficiently large and K_7 and K_8 finite. As before, by application of Kronecker's theorem to Eq. (22), we have that $d_m \rightarrow 0$. The theorem follows directly. It is not hard to show from Eq. (15) that $M(\mathbf{z})$ is minimized for any $\mathbf{z} \epsilon \eta$.

6. An Example

A simple example was devised to compare the first with the second algorithm. The "unknown system" was programmed to be of the form

$$v_m = \sum_{J=0}^{9} \phi_j u_{m-j}, \qquad \phi_j = \exp\left(-\frac{j}{2}\right); j = 0, \cdots, 9 \qquad (23)$$

and the identifier system was programmed to be of the form

$$w_m = \sum_{J=0}^{9} x_j^m u'_{m-j}$$

The relative mean square error was used to measure the performance of the algorithm and, for our example, is defined by

$$\text{rmse} = \frac{\sum_{i=0}^{9} (x_i - \phi_i)^2}{\sum_{i=0}^{9} \phi_i^2} \qquad (24)$$

where the true system values are given by ϕ_i and the estimate of the system by x_i.

The respective elements of the correlation matrix were simulated to satisfy

$$E[u_m u_n] = \delta_{mn}, \qquad E[\epsilon_m \epsilon_n] = K\delta_{mn}, \qquad E[\delta_m \delta_n] = 0 \qquad (25)$$

where u_n and ϵ_n were programmed to simulate independent gaussian random sequences. In Eq. (25), δ_{mn} is the Kronecker delta and K was chosen to be either (a) $K = 1$ or (b) $K = 1/4$, corresponding to an observational signal-to-noise ratio $(\sigma_u^2/\sigma_\epsilon^2)$ of 0 and 6 dB, respectively. Forty thousand samples were used to obtain the system estimates.

Figure 2 illustrates the plot of the rmse as a function of the time index m. As can be seen from Fig. 2, the first algorithm, for both signal-to-noise ratios, approached the theoretical limit defined by the result of Theorem 1 and Eq. (24). Furthermore, the second algorithm continued to decrease, on the average, as m increased, as claimed by Theorem 2.

7. Conclusions

Two sequential identification procedures were presented for the identification of a time invariant, linear system in which no knowledge of the dynamics of the system were known prior to the identification, with the exception that the memory of the system was required to be finite.

The first algorithm for identification required only a mild condition on the covariance function of the input random process and no knowledge of the input or output measurement noise statistics other than that they have finite variances. It was shown that the estimate of the system converged in mean square; however, a bias developed that was due to the input noise and consequently prevented the estimate from being consistent. This error or lack of complete identification, without further knowledge of the statistics of the noises, is an example of what has been called the "structural regression paradox" in the statistical literature.

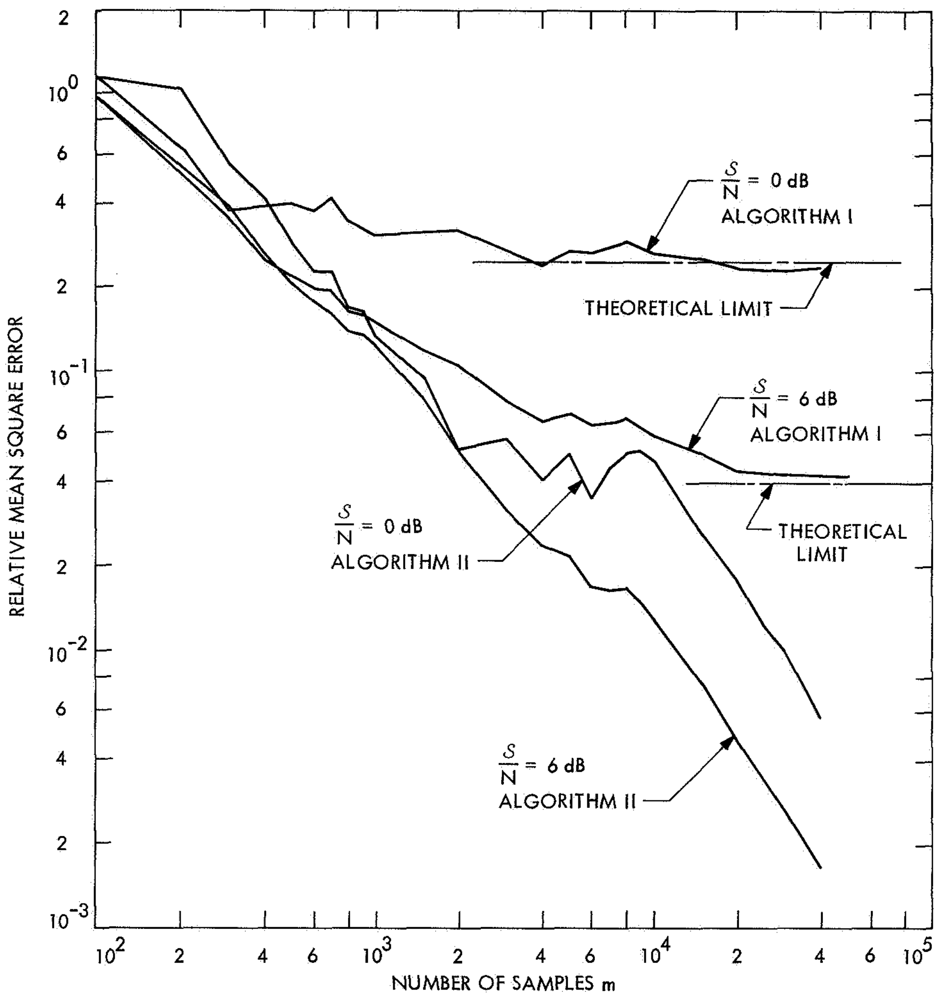

Fig. 2. Relative mean square error as a function of number of samples *m* and theoretical limit of error for first algorithm

The second algorithm was derived based on the additional assumption that the input measurement noise variance was known. With this additional knowledge, it was shown that the estimate of the unknown system converged in mean square to the unknown system. Hence, if the input noise variance is known, this algorithm can be used to obtain a consistent estimate of the unknown system.

Theorem 3 showed that as long as $r = \text{rank } R$ was greater than zero the first algorithm would reduce $M(\mathbf{z})$ to a minimum; however, there could be an uncountable number of values of $\mathbf{z}$ that achieved this minimum.

The computer simulations agreed very well with the theory, indicating that the algorithms are useful and practical methods for identification of linear systems.

References

1. Kushner, H. J., "A Simple Iterative Procedure for the Identification of the Unknown Parameters of a Linear Time Varying Discrete System," *J. Basic Eng.*, pp. 227–235, June 1963.
2. Papers presented at the IFAC Symposium on Identification in Automatic Control Systems, Prague, Czechoslovakia, 1967.
3. Saridis, G. N., and Stein, G., "Stochastic Approximation Algorithms for Linear Discrete-Time System Identification," *National Electronics Conference*, pp. 45–50, 1967.
4. Sakrison, D. J., "Application of Stochastic Approximation Methods to System Optimization," Technical Report 391. Massachusetts Institute of Technology Research Laboratory of Electronics, Cambridge, Mass., 1962.
5. Balakrishnan, A. V., "Determination of Non Linear Systems from Input-Output Data," paper presented at the 54th Meeting of the Princeton University Conference on Identification Problems in Communication and Control Systems, Princeton, N.J., 1963.