

Identifiability of Slowly Varying Systems*

PHILIP H. FISKE

The Analytic Sciences Corporation, Reading, Massachusetts 01867

AND

WILLIAM L. ROOT

*Aerospace Engineering Department, The University of Michigan,
Ann Arbor, Michigan 48109*

A system is conceived of as being slowly varying if it changes slowly enough to permit identification to within a specified error. A generic model is developed to study the identifiability and identification of slowly varying systems. The model is suitable for a large variety of nonlinear, time-varying, causal, bounded memory systems; it has finitely many parameters and is linear in its parameters. Results are obtained with the use of this general model that give guaranteed accuracy of identification as a function of the prior knowledge of the unknown system, the maximum rate of time variation of the system, and the characteristics of output observation noise. To derive these results, a recursive estimation procedure is developed for time-discrete linear dynamical system structures in which the observation noise is statistical but the dynamic equation noise is nonstatistical and is known only to be bounded.

1. INTRODUCTION

Our purpose in this paper is to investigate the identifiability of systems that are changing slowly with time in an unpredictable way. Except in trivial situations, if a system is not time-invariant, a finite-time record of input and output data cannot by itself yield a precise identification of the system, even as it existed during the period the data were taken. The response to the one particular input or sequence of inputs used will be known, but what the response would have been to other inputs cannot be found in general. Even less can be known about the future behavior of the system, unless there is auxiliary information. However, it is reasonable and it is

* Research supported by Air Force Office of Scientific Research, Air Force Systems Command, USAF, under AFOSR Grant Number 72-2328.

common practice to try to identify the unknown systems that are slowly time-varying as if they were time-invariant, updating the identification from time to time in order to track the changes. Such an approach is valid if the system is changing slowly enough to permit sufficiently good approximate identification for the purpose at hand. The problem we address is: How fast can the system change and still permit satisfactory identification from input-output data? The answer to this question depends of course on the criterion of satisfactory identifiability. It also depends on the extent of prior knowledge of the system (e.g., is the system known to be linear?) and on the characteristics and state of prior knowledge of observation noise.

Identifiability can be studied in terms of any system model that is capable of providing a sufficiently faithful representation of the unknown system in question. We choose a particular kind of generic model (i.e., a model with various undetermined parameters) that has the properties of being extremely flexible and of being linear in the parameters. The flexibility of the model makes it applicable to a wide class of systems, nonlinear as well as linear. The linearity in the parameters means that the identification is always a linear estimation problem, which is advantageous when general error formulas are desired that will show the interplay among the factors that control the identifiability. The essential restrictions on generality imposed by the model structure we use are: (1) Input observations must be noise-free, and (2) the unknown system must have finite memory less than or equal to some known bound. Since the theory is an approximation theory, it is not critical in the final interpretation that these conditions be exactly satisfied; however, the mathematical analysis is based on these assumptions.

The generic model is developed in the next two sections. It is based on the notions of classes of systems and their ϵ -representations. In Section 3 a theorem is stated about the identifiability of time-varying systems with noise-free observations of output. This is a relatively simple result which is preparatory to the main identifiability theorem, Theorem 5 of Section 5.

A recursive statistical estimation procedure is developed in Section 4, which is also preparatory to Theorem 5. This procedure applies to a mixed linear model with both stochastic and unknown-but-bounded terms, and is new as far as we are aware. Since the estimation problem is perhaps of some interest in itself, we have treated it a little more generally than is necessary for the application in Section 5.

Some of the results of this paper were presented in the conference papers (Fiske and Root, 1974; Fiske, 1975a); see also Fiske, (1975b), where there is further related material.

2. PRELIMINARIES CONCERNING ϵ -REPRESENTATIONS

The concept of ϵ -representation is introduced and developed somewhat by Root (1971, 1975a); the use of ϵ -representations in identification is discussed abstractly by Root (1973). For the convenience of the reader a summary of those definitions and results concerning ϵ -representations which are needed is provided here.

A class of bounded systems is the 4-tuple $\mathcal{S} = (\mathcal{Y}, f, \mathcal{O}, \mathcal{X})$, where \mathcal{Y} is a Banach space, \mathcal{O} and \mathcal{X} are metric spaces, and f is a continuous map from the topological product $\mathcal{O} \times \mathcal{X}$ into \mathcal{Y} , bounded on \mathcal{X} . \mathcal{Y} , \mathcal{O} , and \mathcal{X} are called the *output space*, *system parameter space*, and *input space*, respectively. If $\alpha \in \mathcal{O}$, $s = (\mathcal{Y}, f, \alpha, \mathcal{X})$ is a *system* belonging to \mathcal{S} . The equation $y = f(\alpha, x)$, $x \in \mathcal{X}$, describes the transformation of inputs to outputs for the system s .

A class of systems \mathcal{S} is *prelinear* if \mathcal{O} is a subset of a linear space and

$$f(c_1\alpha_1 + c_2\alpha_2, x) = c_1f(\alpha_1, x) + c_2f(\alpha_2, x)$$

for all $x \in \mathcal{X}$ whenever the scalars c_1, c_2 and the parameters α_1, α_2 are such that all three terms are defined. Note that the systems in a prelinear class need not have any linearity properties.

Let $\mathcal{F} = \mathcal{F}(\mathcal{X}, \mathcal{Y})$ denote the set of all bounded continuous maps from \mathcal{X} into \mathcal{Y} made into a Banach space in the standard way (see, e.g., Dieudonné, 1960); i.e., if $F \in \mathcal{F}$, $\|F\|$ is defined by $\|F\| = \sup_x \|F(x)\|$, where the norm on the right is the norm in \mathcal{Y} . Define $g: \mathcal{F} \times \mathcal{X} \rightarrow \mathcal{Y}$ by $g(F, x) = F(x)$. Then g is continuous on $\mathcal{F} \times \mathcal{X}$ and actually linear in \mathcal{F} . Thus $\mathcal{S}_{\mathcal{F}} = (\mathcal{Y}, g, \mathcal{F}, \mathcal{X})$ is a linear class of bounded systems.

Note that $f(\alpha, \cdot)$ defines an element, call it H_α , of $\mathcal{F}(\mathcal{X}, \mathcal{Y})$. Let ψ be the map from \mathcal{O} into $\mathcal{F}(\mathcal{X}, \mathcal{Y})$ defined by $\psi(\alpha) = H_\alpha$, and put $\mathcal{H} = \psi(\mathcal{O})$. Then $\mathcal{S}_0 = (\mathcal{Y}, g, \mathcal{H}, \mathcal{X})$ is a class of systems, and is a subclass of $\mathcal{S}_{\mathcal{F}}$. \mathcal{S}_0 is called the *natural representation* of $\mathcal{S} = (\mathcal{Y}, f, \mathcal{O}, \mathcal{X})$ and ψ the *natural mapping* for the class \mathcal{S} . \mathcal{S}_0 is always a prelinear class, whether \mathcal{S} is or not. If \mathcal{S} is itself a prelinear class then it follows that ψ is a restriction of a linear map from the linear span of \mathcal{O} into \mathcal{F} . If ψ is injective and \mathcal{O} and \mathcal{X} are compact metric spaces, then \mathcal{S} is equivalent to its natural representation \mathcal{S}_0 (see Root, 1975a, for definition and details) and nothing is lost by considering \mathcal{S}_0 directly.

Let $\mathcal{S}_1 = (\mathcal{Y}, f_1, \mathcal{O}_1, \mathcal{X})$ be another class of bounded systems with natural mapping ψ_1 . If there exists a map ϕ_1 from \mathcal{H} into \mathcal{O}_1 such that $\|H - \psi_1 \circ \phi_1(H)\| \leq \epsilon$ for all $H \in \mathcal{H}$, (\mathcal{S}_1, ϕ_1) is said to be an ϵ -*representation* of \mathcal{S}_0 (and also to be an ϵ -representation of any class for which \mathcal{S}_0 is the

natural representation). The definition does not require that $\psi_1 \circ \phi_1(H)$ necessarily belong to \mathcal{H} . It does require that \mathcal{S} and \mathcal{S}_1 have the same input and output spaces. An ϵ -representation (\mathcal{S}_1, ϕ_1) of \mathcal{S} is *linear* if \mathcal{S}_1 is a prelinear class and ϕ_1 is a restriction of a linear mapping; it is *finite-dimensional* if \mathcal{O}_1 is a subset of a finite-dimensional Euclidean space; it is *continuous* if both ϕ_1 and ψ_1 are continuous; it is *determined by* $\mathcal{X}_0, \mathcal{X}_0 \subset \mathcal{X}$, if the map ϕ_1 depends only on the functions $H, H \in \mathcal{H}$, restricted to \mathcal{X}_0 . The basic existence theorem is the following.

THEOREM (Root (1975a)). *Let \mathcal{Y} be a Banach space and \mathcal{O} and \mathcal{X} be compact metric spaces. Let $\mathcal{S} = (\mathcal{Y}, f, \mathcal{O}, \mathcal{X})$ be a class of bounded systems. Given $\epsilon > 0$, there exists a continuous, finite-dimensional ϵ -representation (\mathcal{S}_1, ϕ_1) of \mathcal{S} that is determined by a finite subset $\mathcal{X}_0 \subset \mathcal{X}$. \mathcal{S}_1 is a prelinear class. If \mathcal{Y} is a Hilbert space, then in addition ϕ_1 can always be a linear map so that (\mathcal{S}_1, ϕ_1) is a linear ϵ -representation.¹*

The proof of this theorem is by construction and the formulas given by the construction are needed here; so some further discussion is necessary. It is shown that one can find a finite-dimensional subspace \mathcal{Y}' of \mathcal{Y} that has a subset arbitrarily close to the total image set $\mathcal{H}(\mathcal{X}) \subset \mathcal{Y}$. A continuous map π from \mathcal{Y} into \mathcal{Y}' is devised that carries points in $\mathcal{H}(\mathcal{X})$ into sufficiently nearby points of \mathcal{Y}' . In case \mathcal{Y} is a Hilbert space, π can be taken to be the orthogonal projection on \mathcal{Y}' , and this is the only case we consider. Letting $y' = \pi y$, y' is expressed in terms of an arbitrary basis $\{e_1, \dots, e_M\}$ of \mathcal{Y}'

$$y' = \sum_{i=1}^M e_i^*(y') e_i, \tag{2.1}$$

where the e_i^* are continuous linear functionals. Now if the determining set $\mathcal{X}_0 = \{x_1, \dots, x_N\}$, the parameter space \mathcal{O}_1 of the ϵ -representation is a subset of R^{NM} . The map $\phi_1: \mathcal{H} \rightarrow \mathcal{O}_1$ is defined as follows. $\phi_1(H) = h$, where h is that point in R^{NM} with coordinates, with respect to an arbitrary orthonormal basis, given by

$$a_{mn} = e_m^* \circ \pi \circ H(x_n), \quad m = 1, \dots, M; \quad n = 1, \dots, N. \tag{2.2}$$

A set of continuous interpolation functionals $\{\gamma_1(x), \dots, \gamma_N(x)\}$, $x \in \mathcal{X}$, is constructed. These functionals depend on the set $\{x_1, \dots, x_N\}$, and have the properties

¹ It can happen that ϕ_1 is linear when \mathcal{Y} is only a Banach space, but it is not known to us whether this can be guaranteed.

- (1) $\gamma_n(x) \geq 0$, $n = 1, \dots, N$, $x \in \mathcal{X}$,
- (2) $\sum_{n=1}^N \gamma_n(x) = 1$, $x \in \mathcal{X}$,
- (3) $\gamma_n(x_k) = \delta_{nk}$, $n, k = 1, \dots, N$.

Finally, the map f_1 is defined by

$$f_1(h, x) = \left[\sum_{m=1}^M \sum_{n=1}^N \gamma_n(x) a_{mn} \right] e_m. \quad (2.3)$$

An ϵ -representation of the form just described is called a *standard ϵ -representation*. It depends on the set $\{x_1, \dots, x_N\}$, and that set is called its *determining set*.

The integers M and N depend on \mathcal{X} , \mathcal{H} (or \mathcal{O}) and ϵ , but are not uniquely defined simply by the condition that (\mathcal{S}_1, ϕ_1) is a standard ϵ -representation. However, given \mathcal{X} , \mathcal{H} , and ϵ there is a smallest N such that there exists a standard ϵ -representation with a determining set with N elements (we are not concerned with the size of M). With fixed \mathcal{X} and \mathcal{H} this minimum N is nondecreasing as ϵ is made smaller; with fixed ϵ the minimal N is nondecreasing as either \mathcal{X} or \mathcal{H} is made larger in the sense of set inclusion.

Two simple preparatory lemmas are needed that are not given in the references.

LEMMA 1. *For a standard ϵ -representation with \mathcal{Y} a Hilbert space and $\{e_m\}$, $m = 1, \dots, M$, an orthonormal set, ϕ is a restriction of a linear map from $\mathcal{F}(\mathcal{X}, \mathcal{Y}) \rightarrow R^{NM}$ with linear operator norm $\leq N^{1/2}$.*

Proof. Let R_n^M be M -dimensional Euclidean space with basis $\{e_1, \dots, e_M\}$, and let the image of \mathcal{H} given x_n under the mapping defined by (2.2) lie in R_n^M . Let $R^{MN} = R_1^M \oplus \dots \oplus R_N^M$, and assign the basis $\{e_{mn}\}$, $m = 1, \dots, M$; $n = 1, \dots, N$, where e_{mn} is the element of R^{MN} corresponding to e_m in R_n^M . Then with respect to this basis, let $h = \phi_1(H)$, $h' = \phi_1(H')$ have coordinates $\{a_{mn}\}$, $\{a'_{mn}\}$ in R^{NM} , respectively. From (2.2),

$$\begin{aligned} \left\| \sum_{m=1}^M (a_{mn} - a'_{mn}) e_{mn} \right\|^2 &\leq \|H(x_n) - H'(x_n)\|^2 \\ &\leq \|H - H'\|^2, \quad n = 1, \dots, N. \end{aligned}$$

Hence,

$$\|h - h'\|^2 \leq N \|H - H'\|^2$$

from which the assertion follows. ■

LEMMA 2. *With the same hypotheses as for Lemma 1, ψ_1 is a restriction of a linear map from $R^{NM} \rightarrow \mathcal{F}(\mathcal{X}, \mathcal{Y})$ with linear operator norm bounded by one.*

Proof. Because of the proliferation of norms in different spaces which appear, subscripts are used with the norms initially, except for the linear operator norm of ψ_1 which is denoted $|\psi_1|$. Then,

$$\begin{aligned} |\psi_1| &= \sup_{R^{NM}} \frac{\|\psi_1(h)\|}{\|h\|_{R^{NM}}} \\ &= \sup_{R^{NM}} \left\{ \frac{1}{\|h\|} \sup_{\mathcal{X}} \|f_1(h, x)\|_{\mathcal{Y}} \right\} \\ &= \sup_{R^{NM}} \left\{ \frac{1}{\|h\|} \sup_{\mathcal{X}} \left\| \sum_{m=1}^M \left[\sum_{n=1}^N \gamma_n(x) a_{mn} \right] e_m \right\|_{R^M} \right\} \\ &\leq \sup_{R^{NM}} \left\{ \frac{1}{\|h\|} \sup_{\mathcal{X}} \left(\sum_{m=1}^M \left[\sum_{n=1}^N \gamma_n(x) |a_{mn}| \right]^2 \right)^{1/2} \right\} \\ &\leq \sup_{R^{NM}} \left\{ \frac{1}{\|h\|} \sup_{\mathcal{X}} \left(\sum_{m=1}^M \left[\sum_{n=1}^N \gamma_n(x) \cdot \max_n |a_{mn}| \right]^2 \right)^{1/2} \right\} \\ &= \sup_{R^{NM}} \left\{ \frac{1}{\|h\|} \left(\sum_{m=1}^M \max_n |a_{mn}|^2 \right)^{1/2} \right\} \end{aligned}$$

because $\sum_{n=1}^N \gamma_n(x) = 1$ for all $x \in \mathcal{X}$. Then

$$|\psi_1| \leq \sup_{R^{NM}} \left[\frac{\sum_{m=1}^M \max_n |a_{mn}|^2}{\sum_{m=1}^M \sum_{n=1}^N a_{mn}^2} \right]^{1/2} \leq 1. \quad \blacksquare$$

3. A MODEL FOR TIME-VARYING SYSTEMS

Some further notations, conventions, and definitions are useful. In what follows we deal with spaces of functions (or equivalence classes of functions) on R^1 or subsets of R^1 ; in each case the functions take values in a finite-dimensional Euclidean space. $L^2(a, b)$ denotes the L^2 -space formed from functions defined on (a, b) and square integrable Lebesgue. We shall identify the spaces $L^2(a, b)$ and $L^2(t + a, t + b)$, t any real number, without further comment. The symbol \mathcal{A} will sometimes be used as a generic symbol for

an interval in R^1 (of the form $(a, b]$, if the endpoints matter). The linear operators L_a , P_a and P_Δ are defined on functions x on R^1 by

$$\begin{aligned} (L_a x)(t) &= x(t + a), \\ (P_a x)(t) &= x(t), \quad t \leq a, \\ &= 0, \quad t > a, \\ (P_\Delta x)(t) &= x(t), \quad t \in \Delta, \\ &= 0, \quad t \notin \Delta. \end{aligned}$$

The symbol \mathcal{X}_e denotes a set of functions x on R^1 that satisfies the conditions

- (i) (shift invariance) $x \in \mathcal{X}_e \Rightarrow L_a x \in \mathcal{X}_e$ for all $a \in R^1$;
- (ii) (projection property) $x \in \mathcal{X}_e \Rightarrow P_a x \in \mathcal{X}_e$, $(I - P_a)x \in \mathcal{X}_e$ and $P_\Delta x \in \mathcal{X}_e$ for all a and Δ .

To avoid confusion between different uses of the word system, we introduce the term *extended system* to represent the kind of ongoing, in general time-varying, system that is to be identified. Precisely, an extended system is a triple $(\mathcal{Y}_e, H^e, \mathcal{X}_e)$ where \mathcal{X}_e is as just defined, \mathcal{Y}_e is a linear function space on R^1 that also satisfies

- (iii) shift invariance, and
- (iv) projection property.

H^e is a map² from \mathcal{X}_e into \mathcal{Y}_e . The extended system is causal if $P_a H^e(x) = P_a H^e(P_a x)$ for all $x \in \mathcal{X}_e$ and all a . It has *bounded memory* m if $(I - P_a) H^e(x) = (I - P_a) H^e[(I - P_{a-m})(x)]$ for all $x \in \mathcal{X}_e$ and all a . These can be combined: An extended system is causal with bounded memory m if and only if for every $T > 0$

- (v) $(P_{t+T} - P_t) H^e(x) = (P_{t+T} - P_t) H^e[(P_{t+T} - P_{t-m})(x)]$ for all t and all $x \in \mathcal{X}_e$ (see Root, 1975b).

The idea we follow in modeling a natural system operating for an indefinite time period is first to represent it as an extended system with certain additional restrictions on \mathcal{X}_e , \mathcal{Y}_e and H^e , and then to regard the extended system as corresponding to a trajectory in some fixed class of systems, as defined in Section 1. We choose to set up the problem so that the topological function

² In Root (1975b), spaces satisfying the conditions for \mathcal{X}_e and \mathcal{Y}_e are topologized; the H^e are taken to be continuous maps, and classes of what are here called extended systems are formed. A much more elaborate structure is developed than is necessary for our purposes here.

spaces that appear (this does not include \mathcal{X}_e and \mathcal{Y}_e) are L^2 -spaces. This is not necessary, but it seems sufficiently general and it facilitates the statistical estimation discussed in the next section.

Let T and m be fixed positive numbers. T will be the duration of each observation interval; m will be a bound on the memory duration. The number m is chosen to fit the problem; either the natural systems in question have bounded memory m , or they can be approximated sufficiently well by having the memory truncated to m . T is arbitrary. Let $\mathcal{S} = (\mathcal{Y}, f, \mathcal{O}, \mathcal{X})$ be a class of systems satisfying the conditions

- (a) \mathcal{X} is a compact subset of $L^2(-m, T)$,
- (b) \mathcal{Y} is $L^2(0, T)$,
- (c) \mathcal{O} is compact.

We deal with the natural representation of \mathcal{S} , $\mathcal{S}_0 = (\mathcal{Y}, g, \mathcal{H}, \mathcal{X})$. It follows, (c'), that \mathcal{H} is a compact subset of $\mathcal{F}(\mathcal{X}, \mathcal{Y})$.

Now let $(\mathcal{Y}_e, H^e, \mathcal{X}_e)$ be an extended system that satisfies the following additional conditions. The input space \mathcal{X}_e satisfies

- (vi) $P_\Delta \mathcal{X}_e \subset L^2(\Delta)$ for any Δ ,
- (vii) $P_{(t-m, t+T)} \mathcal{X}_e = L_{-t} \mathcal{X}$.

Note that, by virtue of (i), it is sufficient that (vi) hold for some Δ and that (vii) hold for some t . The output space \mathcal{Y}_e satisfies

- (viii) $P_\Delta \mathcal{Y}_e \subset L^2(\Delta)$ for any Δ .

Given any t , the map H^e induces a map H_t from $P_{(t-m, t+T)} \mathcal{X}_e$ into $L^2(t, t+T)$ defined by

$$H_t(x_t) = P_{(t, t+T)} H^e(P_{(t-m, t+T)} x), \quad x \in \mathcal{X}_e,$$

$$x_t = P_{(t-m, t+T)} x.$$

In fact, because of (v) the first equation can be written $H_t(x_t) = P_{(t, t+T)} H^e(x)$. Since by (vii) all $x \in \mathcal{X}$ are translates of such x_t , H_t can be regarded as a map from \mathcal{X} into $\mathcal{Y} = L^2(0, T)$. It is required that H^e be such that

- (ix) $H_t \in \mathcal{H}$.

With \mathcal{S}_0 as given it is clear that any extended system that satisfies the additional conditions (v), ..., (ix) generates a trajectory $\{H_t\}$, $t \in R^1$, in the subset \mathcal{H} of $\mathcal{F}(\mathcal{X}, \mathcal{Y})$. We regard an extended system satisfying these conditions as "identifiable" if it is possible to determine H_{t_0} for any t_0 on the basis of input-output data taken over a finite time interval terminating

at time $t_0 + T$. We regard the extended system as “approximately identifiable” if it is possible to determine H_{t_0} to within a prescribed tolerance, in a prescribed sense, under the same conditions. As indicated in the Introduction, we are concerned with approximate identifiability.

To make a *measurement* on the extended system at t will mean to apply an input signal during the time interval $(t - m, t + T]$ and observe the output during the time interval $(t, t + T]$. Measurements may be made for overlapping time intervals, but if they are, the successive inputs are necessarily related. To allow a completely arbitrary choice of successive inputs, we suppose that measurements are made at the times $t_1 = 0$, $t_2 = m + T, \dots, t_{k+1} = k(m + T), \dots$. The maps H_{t_k} , denoted H_k , describe a discrete trajectory in \mathcal{H} . Setting $W_k = H_{k+1} - H_k$ gives the equations

$$\begin{aligned} H_{k+1} &= H_k + W_k, \\ y_k &= H_k(\bar{x}_k), \quad k = 1, 2, \dots, \end{aligned} \tag{3.1}$$

to describe the successive measurements, where \bar{x}_k is the input signal applied during the interval $(t_k - m, t_k + T]$ and y_k is the output observed during the interval $(t_k, t_k + T]$.

It is to be noted that even though we are defining a system to be an input-output map, there are no serious restrictions on the “state” of the extended system at any time during the measurement process. In particular, since the duration of the input interval is equal to that of the output interval plus a bound on the duration of the system memory, the “output” for each measurement depends only on the “input” for that measurement, and is independent of the state of the system at the beginning of the measurement. This holds true even in the case where measurements are made for overlapping time intervals.

To get a finite-parameter model for the H_k we employ ϵ -representations. Let $\epsilon > 0$ be fixed at an arbitrary value; then by virtue of conditions (a), (b), (c) on \mathcal{S} there exists a standard ϵ -representation (\mathcal{S}_1, ϕ_1) as given by Eqs. (2.2) and (2.3), with determining set $\{x_1, \dots, x_N\}$. As before $\mathcal{S}_1 = (\mathcal{Y}, f_1, \mathcal{O}_1, \mathcal{X})$, and the parameter set $\mathcal{O}_1 \subset R^{NM}$. Also, π is the orthogonal projection from \mathcal{Y} onto \mathcal{Y}' , which is an M -dimensional subspace; $\{e_1, \dots, e_M\}$ is an orthogonal basis for \mathcal{Y}' , and the orthogonal basis for R^{NM} is chosen as in Lemma 1.

The fundamental identification problem is to find H_k , $k = 1, 2, \dots$; we are satisfied if we can find a suitably close estimate of $h_k = \phi_1 H_k$, the corresponding vector parameter for the standard ϵ -representation. Suppose, temporarily, that the extended system is time-invariant so that

$H_k = H_1$ for all k . Let \bar{x}_k , $k = 1, \dots, N$, be chosen to be the N elements of the determining set, thus $\bar{x}_k = x_k$, $k = 1, \dots, N$. Then, from (2.2)

$$\begin{aligned} y_k' &= \pi(y_k) = \pi \circ H_1(x_k) \\ &= \sum_{m=1}^M (\pi \circ H_1(x_k), e_m) e_m = \sum_{m=1}^M a_{mk} e_m, \quad k = 1, \dots, N. \end{aligned} \quad (3.2)$$

Let h_1 be represented by the column vector

$$h_1 = [a_{11}, \dots, a_{M1}; a_{12}, \dots, a_{M2}; \dots; a_{1N}, \dots, a_{MN}]^T;$$

then (3.2) can be written as a matrix equation

$$y_n' = X_n h_1, \quad n = 1, \dots, N, \quad (3.3)$$

where y_n' is now interpreted as a column vector of length M , and X_n is an $M \times MN$ matrix which in partitioned form is $X_n = [0 \vdots \dots \vdots 0 \vdots I \vdots 0 \vdots \dots \vdots 0]$, each block being $M \times M$ and the identity matrix appearing in the n th block. Setting

$$y^1 = \begin{bmatrix} y_1' \\ \dots \\ \vdots \\ \dots \\ y^{N'} \end{bmatrix}$$

and combining Eqs. (3.3) yields

$$y^1 = X h_1 = h_1,$$

since

$$X = \begin{bmatrix} X_1 \\ \dots \\ \vdots \\ \dots \\ X_N \end{bmatrix} = I.$$

Thus, in this case of noise-free identification of a time-invariant extended system, an identification to within ϵ is obtained by making measurements with each input signal in the determining set. The parameter values a_{mn} are actually the coordinates of the projections of the outputs on an appropriate M -dimensional subspace with a certain orthonormal basis.

Now suppose that the extended system is continually changing so that each measurement is performed on a (slightly) different system. From (3.1) and the argument leading to (3.3) we have

$$\begin{aligned} h_{k+1} &= h_k + w_k, \\ y'_k &= X_k h_k, \quad k = 1, 2, \dots, N, \end{aligned} \quad (3.5)$$

as the equations describing the situation at the k th measurement in terms of the standard ϵ -representation. Since ϕ_1 is linear, $w_k = \phi_1 W_k$. We want to consider blocks of N measurements, and for the purposes of the next section, successive blocks of N measurements, each with the same successive inputs from the determining set $\{x_1, \dots, x_N\}$. Equations (3.5) are then meaningful for all $k = 1, 2, \dots$ if $X_k = X_j$, $j = k - [k/N]N$, where $[a]$ denotes the integer part of a . Put

$$h^n = h_{NN}, \quad y^n = \begin{bmatrix} y'_{(n-1)N+1} \\ y'_{(n-1)N+2} \\ \vdots \\ y'_{nN} \end{bmatrix}, \quad (3.6)$$

$$\epsilon^n = \begin{bmatrix} -X_1(w_{(n-1)N+1} + w_{(n-1)N+2} + \dots + w_{nN-1}) \\ -X_2(w_{(n-1)N+2} + \dots + w_{nN-1}) \\ \vdots \\ -X_{N-1}w_{nN-1} \\ 0 \end{bmatrix},$$

and

$$w^n = \sum_{k=nN}^{(n+1)N-1} w_k.$$

Then the successive blocks of N equations given by (3.5) can be written

$$\begin{aligned} h^{n+1} &= h^n + w^n, \\ y^n &= Xh^n + \epsilon^n = h^n + \epsilon^n, \quad n = 1, 2, \dots, \end{aligned} \quad (3.7)$$

since X is the $NM \times NM$ identity matrix. With the w_n , and hence the w^n and ϵ^n , unknown, it is now impossible to find any of the h^n precisely from observations of the y^n ; however, if each w^n is small one would expect $h^n = y^n$ to be a pretty good approximation to h^n . We have the following simple result.

LEMMA 3. *In the model (3.7) we let w_n satisfy only the condition $\|w_n\| \leq \eta$, where $\|w_n\|$ is the Euclidean norm of w_n in R^{NM} . Then $\bar{h}^n = y^n$ satisfies $\|\bar{h}^n - h^n\| \leq (N - 1)\eta$, and no estimate of h^n formed from y^n can have guaranteed error less than $(N - 1)\eta$.*

Proof. It is sufficient to consider only y^1 in (3.7), that is, to consider $k = 1, \dots, N$, in (3.5). Given the same h_1 , the two following sets of values for the w_k yield the same observations y_k' . (1) $w_1 = w_2 = \dots = w_{N-1}$, $\|w_1\| = \eta$, $w_1 \in \mathcal{N}^\perp(X_1)$ (the orthogonal complement of the null space of X_1); (2) $\tilde{w}_1 = \tilde{w}_2 = \dots = \tilde{w}_{N-1} = -w_1$. In fact, in either case $y_k' = X_k h_1$, $k = 1, \dots, N$, since all the w_k and \tilde{w}_k belong to $\mathcal{N}(X_k)$ for $k > 1$. The values of h^n for the two different cases differ by an element of norm $2(N - 1)\eta$. This proves the second assertion. It is easy to see that the choice of w_1, \dots, w_{N-1} just given defines a worst case for \bar{h}^n , and the first assertion follows immediately. ■

The following theorem summarizes the discussion of the model (3.7) by stating to what extent identifiability of a time-varying system, meeting certain conditions, can be guaranteed. This is, of course, a statement of the situation when noise-free observations of the output are possible. It is preliminary to a corresponding result in the final section that takes account of noisy observations of output.

THEOREM 1. *Let $(\mathcal{Y}_e, H^e, \mathcal{X}_e)$ be an extended system satisfying (i), ..., (viii) for some $m > 0$. Suppose there is a class of systems \mathcal{S} satisfying (a), (b), (c) with the same m and arbitrary $T > 0$ and such that the extended system satisfies (ix) with respect to \mathcal{S} . Finally let it be required that the changes in the extended system during one measurement interval are always such that $\|H_{k+1} - H_k\| \leq \delta$ for all k .*

Fix $\epsilon > 0$ and let N be the minimum number of elements in the determining set of a standard ϵ -representation of \mathcal{S} . Then there exists an estimate \bar{H}_k of H_k , for any k , formed from a preceding block of N measurements such that

$$\|\bar{H}_k - H_k\| \leq \epsilon + (N - 1) N^{1/2} \delta.$$

Proof. Consider the model (3.7) based on the standard ϵ -representation of \mathcal{S} . From Lemma 1 and the condition that $\|H_{k+1} - H_k\| \leq \delta$ for all k ,

$$\|w_k\| \leq \|\phi_1\| \|W_k\| \leq N^{1/2} \delta.$$

Then from Lemma 3 the estimate $\bar{h}^n = y^n$ satisfies $\|\bar{h}^n - h^n\| \leq (N - 1) N^{1/2} \delta$. Put $\bar{H}^n = \psi_1(\bar{h}^n)$; then from Lemma 2,

$$\|\bar{H}^n - \psi_1(h^n)\| \leq (N - 1) N^{1/2} \delta.$$

Since $h^n = \phi_1(H_{nN})$, and (\mathcal{S}_1, ϕ_1) is an ϵ -representation, putting $\bar{H}_{nN} = \bar{H}^n$ yields

$$\|\bar{H}_{nN} - H_{nN}\| \leq \epsilon + (N-1)N^{1/2}\delta,$$

from the triangle inequality. However, since a block of measurements may be started anywhere, nN may actually take on any integer value, which proves the result. ■

Remarks. (1) Conditions (i), (ii), (vi) on \mathcal{X}_ϵ , together with the compactness of \mathcal{X} , are compatible and not even very restrictive (see Proposition 2.4 of Root, 1975b). The compactness condition of input spaces may seem confining to one used to dealing with linear system theory, but in a practical situation it can often be argued that potential inputs must indeed belong to a compact set because of very real physical constraints (see comments in Root, 1971).

(2) It is evident that even in the case of noise-free observations the identifiability of a time-varying system depends on a trade-off between actual rate of system variation on the one hand and the state of prior knowledge (specification of \mathcal{O}) and size of the class of admissible inputs (specification of \mathcal{X}) on the other. This is indicated in Theorem 1 in the dependence of N on ϵ , \mathcal{O} and \mathcal{X} . It would be highly desirable to estimate the dependence of N on ϵ for simple examples of \mathcal{X} and \mathcal{O} , but no attempt is made to do this here. The problem is akin to estimating the ϵ -entropy of a set, but is more complicated because of the sense in which the x_n are required to cover \mathcal{X} .

(3) In many situations some simple ad hoc construction can be used to obtain a linear-in-the-parameters model of the form of Eq. (3.7) (but usually not with $X = I$), and the statistical estimation theory of the next section will apply to such models.

We now suppose that the observations of the output of the extended system are contaminated by additive noise. Then Eqs. (3.1) are replaced by

$$\begin{aligned} H_{k+1} &= H_k + W_k, \\ z_k &= H_k(\bar{x}_k) + V_k, \end{aligned} \tag{3.9}$$

where V_k is a random variable³ taking values in $L^2(0, T)$. Assume

³ That is, we take V_k to be a strongly measurable map from a complete probability space into $L^2(0, T)$ (see, e.g., Barucha-Reid, 1972).

(α) $E(V_k)$ exists and is equal to zero.

(β) V_k has a covariance operator Γ (which will necessarily be self-adjoint and of trace class).

(γ) V_k and V_j are uncorrelated for $k \neq j$; i.e., the cross-covariance operator is zero.

The development leading to (3.7) can now be repeated with the only change being that in each observation equation there is an additive noise term. Let z_k' be defined to be $\pi(z_k)$. Then, since π is linear, Eqs. (3.5) are replaced by

$$\begin{aligned} h_{k+1} &= h_k + w_k, \\ z_k' &= X_k h_k + v_k, \end{aligned} \quad (3.10)$$

where $v_k = \pi V_k$ is represented as a column vector with respect to the basis $\{e_1, \dots, e_M\}$. The random vectors v_k have mean zero, are uncorrelated and have covariance operator $\pi \Gamma \pi$, which will be represented by the covariance matrix R . Using the same convention as before as regards superscripts versus subscripts leads to

$$\begin{aligned} h^{n+1} &= h^n + w^n, \\ z^n &= h^n + \epsilon^n + v^n, \end{aligned} \quad (3.11)$$

where v^n is a random vector with mean zero and covariance matrix $\tilde{R} = N$ -block diagonal $[R : R : \dots : R]$. Furthermore, v^n and v^m are uncorrelated for $n \neq m$.

Equations (3.11) describe the model we consider for identification of a time-varying system in the presence of output noise. In the noise-free case one can compute $\hat{h}^n = h^n + \epsilon^n$ as an estimate of h^n ; in the noisy case one can make a statistical estimate \hat{h}^n of $h^n + \epsilon^n$, which is again regarded as an estimate of h^n . In both cases there is nonremovable error due to the presence of ϵ^n . In the noise-free case one block of N measurements suffices; in the noisy case continuing measurements allow for a reduction of statistical error, as usual. In the next section the statistical estimation problem is treated.

4. PARAMETER ESTIMATION IN THE TIME-VARYING SYSTEM MODEL

The estimation problem posed by (3.11) is a "mixed" problem which is somewhat unconventional; the terms w^n and ϵ^n are nonstochastic disturbances while the term v^n is stochastic. Even if the ϵ^n were zero, the

usual Kalman recursive linear least-mean-squared (LMS) error estimation theory would not apply because of the nature of w^n . Since such an estimation problem is at least of some academic interest,⁴ we generalize it a little, considering however only the case $\epsilon^n = 0$, before constructing an estimate. This more general version also is of use for some time-varying system models formulated on an ad hoc basis with no reference to standard ϵ -representations, as mentioned above. When the solution we obtain for the estimation problem with $\epsilon^n = 0$ is applied to (3.11) an adjustment has to be made to account for the presence of ϵ^n and the extra error it causes, and this is done in the next section.

The criterion used for the quality of estimates is, as usual, mean-squared error. However, in the model used the mean-squared error depends on the values of the unknown, bounded, but nonstochastic quantities. What we would like to do is minimize the maximum mean-squared error. However, the recursive linear estimate constructed only minimizes an upper bound on the maximum mean-squared error, so a comparison lower bound is obtained to provide a guarantee that the estimation procedure is not far from optimum. No linear estimation procedure can yield an estimate with mean-squared error less than the number given by the lower bound for all possible values of the parameters.

We consider

$$\begin{aligned} h^{n+1} &= h^n + w^n, \\ z^n &= Xh^n + v^n, \quad n = 1, 2, \dots, \end{aligned} \tag{4.1}$$

where again h^n is the vector parameter to be estimated, but where the assumptions are now: h^n is a p -dimensional vector; z^n is an r -dimensional vector, $r \geq p$; X is arbitrary except that $\mathcal{N}(X) = 0$;⁵ v^n is a random vector with first and second moments satisfying

$$\begin{aligned} Ev^n &= 0, \\ Ev^n(v^m)^T &= R\delta_{nm}, \end{aligned} \tag{4.2}$$

⁴ Recursive estimation for models where the noise is unknown-but-bounded has been studied by Schweppe, Schlaepfer, Bertsekas and Rhodes, and others (see Schweppe, 1974, and other references given there). The estimator developed here is different from any previously discussed as far as we know.

⁵ That part of h that lies in the null space of X is not estimable anyway, so there is no real loss in generality in cutting down the parameter space, if necessary, to coincide with $\mathcal{N}^\perp(X)$.

with the covariance matrix R strictly positive definite, and w^n is bounded componentwise, as will be specified. Let $\{\psi_i\}$, $i = 1, \dots, p$, be a set of orthonormal eigenvectors belonging to $(X^T R^{-1} X)^{-1}$, i.e.,

$$(X^T R^{-1} X)^{-1} \psi_i = \sigma_i^2 \psi_i, \quad i = 1, \dots, p. \quad (4.3)$$

Then it is required that⁶

$$|\psi_i^T w^n| = |(w^n, \psi_i)| \leq \eta_i, \quad i = 1, \dots, p; \text{ all } n. \quad (4.4)$$

Note that since $\mathcal{N}(X) = 0$ and R is strictly definite, $(X^T R^{-1} X)^{-1}$ exists and the eigenvalues $\{\sigma_i^2\}$ are real and positive.

A recursive estimator for h^n is now developed. It is suggested by the nonrecursive modified linear unbiased minimum-variance (modified LUMV) estimator described by Root (1973); there is also a vague resemblance to the standard Kalman recursive estimator. To start, let \hat{h}^1 be the ordinary LUMV estimate of h^1 using the observation z^1 , i.e.,

$$\hat{h}^1 = (X^T R^{-1} X)^{-1} X^T R^{-1} z^1 = C z^1, \quad (4.5)$$

where the matrix C is defined implicitly by (4.5). As is well known (Albert, 1972), the mean-squared error in this estimate satisfies

$$\begin{aligned} E \|\hat{h}^1 - h^1\|^2 &= E \|C(Xh^1 + v^1) - h^1\|^2 \\ &= E \|Cv^1\|^2 = \text{Tr}[(X^T R^{-1} X)^{-1}] \\ &= \sum_{i=1}^p \sigma_i^2. \end{aligned} \quad (4.6)$$

Here and in the rest of this section the norm is the Euclidean norm in R^p . It is convenient to introduce the notation $b_{1i}^2 = \sigma_i^2$, $i = 1, \dots, p$, so that

$$E \|\hat{h}^1 - h^1\|^2 = \sum_{i=1}^p b_{1i}^2. \quad (4.7)$$

Now consider the second observation

$$z^2 = Xh^2 + v^2 = Xh^1 + Xw^1 + v^2,$$

and rewrite this as

$$z^2 - X\hat{h}^1 = X[(h^1 - \hat{h}^1) + w^1] + v^2. \quad (4.8)$$

⁶ We occasionally use inner-product space notation instead of vector-matrix notation in this section.

Define $\theta^2 \stackrel{d}{=} z^2 - X\hat{h}^1$ and $\xi^1 \stackrel{d}{=} (h^1 - \hat{h}^1) + w^1 = h^2 - \hat{h}^1$ so that (4.8) becomes

$$\theta^2 = X\xi^1 + v^2. \quad (4.9)$$

The LUMV estimate of ξ^1 is

$$C\theta^2 = \xi^1 + Cv^2. \quad (4.10)$$

Equation (4.10) can be interpreted as describing a linear model for estimation in which the linear transformation is the identity, the "observation" is $C\theta^2$ and the "noise" is Cv^2 . Since the LUMV estimate of ξ^1 in (4.10) is just the observation itself, there is no conflict in thinking of (4.10) this way. We now wish to use the a priori information that we have about ξ^1 , i.e., the information in (4.4) and (4.7), to modify the LUMV estimate of ξ^1 so as to obtain a new estimate with smaller mean-squared error.

Consider a completely arbitrary linear estimate $\tilde{\xi}^1$ of ξ^1 in (4.10) and expand it in terms of the $\{\psi_i\}$. Thus,

$$\tilde{\xi}^1 = \sum_{i,j=1}^p a_{ij}(C\theta^2, \psi_j) \psi_i, \quad (4.11)$$

where the $\{a_{ij}\}$ are real numbers. The error in this estimate is

$$\tilde{\xi}^1 - \xi^1 = \sum_{i=1}^p \left[(a_{ii} - 1)(\xi^1, \psi_i) + \sum_{\substack{j=1 \\ (j \neq i)}}^p a_{ij}(\xi^1, \psi_j) + \sum_{j=1}^p a_{ij}(Cv^2, \psi_j) \right] \psi_i. \quad (4.12)$$

Then, from (4.2) and (4.3), and using the facts that ξ^1 is uncorrelated with v^2 and that $CRC^T = (X^T R^{-1} X)^{-1}$, it follows that

$$E \|\tilde{\xi}^1 - \xi^1\|^2 = \sum_{i=1}^p E \left[(a_{ii} - 1)(\xi^1, \psi_i) + \sum_{\substack{j=1 \\ (j \neq i)}}^p a_{ij}(\xi^1, \psi_j) \right]^2 + \sum_{i,j=1}^p a_{ij}^2 \sigma_j^2. \quad (4.13)$$

We should like to choose the $\{a_{ij}\}$ so as to minimize the supremum of $E \|\tilde{\xi}^1 - \xi^1\|^2$ for all h and all w satisfying (4.4). However this minimization problem appears to be extremely complicated, and perhaps not even worthwhile in view of what follows, so we content ourselves with minimizing the obvious upper bound to (4.13) given by

$$E \|\tilde{\xi}^1 - \xi^1\|^2 \leq 2 \sum_{i=1}^p (a_{ii} - 1)^2 E[(\xi^1, \psi_i)]^2 + 2 \sum_{i=1}^p E \left[\sum_{\substack{j=1 \\ (j \neq i)}}^p a_{ij}(\xi^1, \psi_j) \right]^2 + \sum_{i,j=1}^p a_{ij}^2 \sigma_j^2. \quad (4.14)$$

To minimize the right side of (4.14), one should obviously set $a_{ij} = 0$ for $i \neq j$. Making this assignment reduces (4.13) to

$$E \|\tilde{\xi}^1 - \xi^1\|^2 = \sum_{i=1}^p (a_{ii} - 1)^2 E[(\xi^1, \psi_i)]^2 + \sum_{i=1}^p a_{ii}^2 \sigma_i^2. \quad (4.15)$$

Now

$$\begin{aligned} E[(\xi^1, \psi_i)]^2 &= E[(h^1 - \hat{h}^1 + w^1, \psi_i)]^2 \\ &= E[(Cv^1, \psi_i)^2] - 2(w^1, \psi_i) E[(Cv^1, \psi_i)] + E[(w^1, \psi_i)^2] \\ &\leq \sigma_i^2 + 2\eta_i \sigma_i + \eta_i^2 = (b_{1i} + \eta_i)^2, \end{aligned} \quad (4.16)$$

so

$$E \|\tilde{\xi}^1 - \xi^1\|^2 \leq \sum_{i=1}^p [(a_{ii} - 1)^2 (b_{1i} + \eta_i)^2 + a_{ii}^2 \sigma_i^2]. \quad (4.17)$$

Obviously the bound in (4.16) could be improved to $b_{1i}^2 + \eta_i^2$, since the unbiasedness of \hat{h}^1 causes the cross term to vanish. However, this calculation is used all over again in an inductive proof, and in general the preceding estimate \hat{h}^n is biased so the use of the Schwarz inequality is necessary.

The upper bound in (4.17) is minimized by putting

$$a_{ii} = \frac{(b_{1i} + \eta_i)^2}{(b_{1i} + \eta_i)^2 + \sigma_i^2}, \quad i = 1, \dots, p. \quad (4.18)$$

We let $\hat{\xi}^1$ denote the estimate of ξ^1 given by (4.11) with a_{ii} as given by (4.18) and $a_{ij} = 0$, $i \neq j$. Since $h^2 = \hat{h}^1 + \xi^1$, we take as our estimate of h^2 ,

$$\hat{h}^2 \stackrel{a}{=} \hat{h}^1 + \hat{\xi}^1. \quad (4.19)$$

It is perhaps interesting to note that the definition of \hat{h}^2 in (4.19) is analogous to a key step in a standard derivation of the Kalman recursive least-mean-squared error estimate; the difference is that there the step is shown to lead to an optimum estimate by a projection argument, whereas here there is no such structure available. With the use of (4.17) a short calculation shows that the error in \hat{h}^2 satisfies the following inequality,

$$\begin{aligned} E \|\hat{h}^2 - h^2\|^2 &= E \|(\hat{h}^1 + \hat{\xi}^1) - (\hat{h}^1 + \xi^1)\|^2 \\ &= E \|\hat{\xi}^1 - \xi^1\|^2 \\ &\leq \sum_{i=1}^p \frac{(b_{1i} + \eta_i)^2 \sigma_i^2}{(b_{1i} + \eta_i)^2 + \sigma_i^2} = \sum_{i=1}^p \frac{(\sigma_i + \eta_i)^2 \sigma_i^2}{(\sigma_i + \eta_i)^2 + \sigma_i^2} \\ &= \sum_{i=1}^p b_{2i}^2, \end{aligned} \quad (4.20)$$

where

$$b_{2i}^2 \stackrel{d}{=} \frac{(b_{1i} + \eta_i)^2 \sigma_i^2}{(b_{1i} + \eta_i)^2 + \sigma_i^2}.$$

Repetition of the procedure just described leads to the recursive estimate

$$\begin{aligned} \hat{h}^n &= \hat{h}^{n-1} + \hat{\xi}^{n-1} \\ \hat{\xi}^{n-1} &= \sum_{i=1}^p \frac{(b_{(n-1)i} + \eta_i)^2}{(b_{(n-1)i} + \eta_i)^2 + \sigma_i^2} (Cz^n - \hat{h}^{n-1}, \psi_i) \psi_i \\ b_{ni}^2 &= \frac{(b_{(n-1)i} + \eta_i)^2 \sigma_i^2}{(b_{(n-1)i} + \eta_i)^2 + \sigma_i^2}, \quad i = 1, \dots, p \end{aligned} \tag{4.21}$$

with initial conditions given by

$$\begin{aligned} \hat{h}^1 &= Cz^1, \\ b_{1i}^2 &= \sigma_i^2, \quad i = 1, \dots, p. \end{aligned} \tag{4.22}$$

THEOREM 2. *Given the statistical model defined by Eqs. (4.1) and subject to the conditions of (4.2) and (4.4) and to $\mathcal{N}(X) = 0$. The recursive estimates defined by Eqs. (4.21) and (4.22) satisfy the error bound*

$$E \| \hat{h}^n - h^n \|^2 \leq \sum_{i=1}^p b_{ni}^2. \tag{4.23}$$

Proof. At the k th stage the model (4.1) gives the observation equation

$$z^k = Xh^k + v^k = Xh^{k-1} + Xw^{k-1} + v^k,$$

which can be rewritten as

$$\theta^k = X\xi^{k-1} + v^k,$$

where $\theta^k \stackrel{d}{=} z^k - Xh^{k-1}$ and

$$\xi^{k-1} \stackrel{d}{=} (h^{k-1} - \hat{h}^{k-1}) + w^{k-1} = h^k - \hat{h}^{k-1}.$$

From these relations Eqs. (4.21) follow for $n = k$ exactly as they did for $n = 2$ and the calculation of the accompanying error bound (4.23) also goes the same way. The assertion follows by induction. ■

The behavior of the error bound in (4.23) as $n \rightarrow \infty$ is of concern. Before presenting a result for this limiting case we state without proof a simple preparatory lemma (see Fiske, 1975b).

LEMMA 4. *If f is a real-valued C^2 function defined on the positive half-line and satisfies*

- (i) $\lim_{x \rightarrow 0^+} f(x) > 0$,
- (ii) $\lim_{x \rightarrow \infty} f(x) = \text{constant} < \infty$,
- (iii) $f'(x) > 0, \quad x > 0$,
- (iv) $f''(x) < 0, \quad x > 0$,

then,

- (a) $f(x) = x$ has a unique solution x^* for $x > 0$,
- (b) the sequence $x_n = f(x_{n-1}), x_1 > 0$, converges to x^* .

THEOREM 3. *The equation*

$$\frac{(x^{1/2} + \eta_i)^2 \sigma_i^2}{(x^{1/2} + \eta_i)^2 + \sigma_i^2} = x, \quad x > 0, \tag{4.24}$$

with $\eta_i, \sigma_i > 0$, has a unique solution x_i^* . For the recursive estimates of Theorem 2, $\lim_{n \rightarrow \infty} b_{ni}^2 = x_i^*$.

Proof. Follows from Lemma 4 and (4.21). ■

A graph of x_i^* is shown for certain values of η_i and σ_i in Fig. 1.

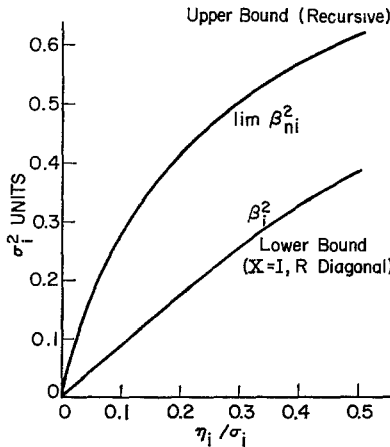


FIG. 1. Upper and lower bounds on MSE (i th component).

EXAMPLE. For a system with input-output relation characterizable by a linear integral operator, or more generally by a polynomial integral operator (Volterra polynomial), there is an easy and obvious way to get a linear-in-the-parameters model of the form of (4.1) to which Theorem 2 applies. We indicate how this goes for a second-degree polynomial, and then consider details in a simple special case. It will be clear how this procedure generalizes to higher degree polynomials. Two comments should be made, even though they are obvious. First, some approximation is necessary to reduce the model to a finite-dimensional one; second, this is an example of an ad hoc system model and has nothing to do directly with ϵ -representations.

Suppose the system is described by

$$y(t) = \int_{\tau-m}^{\tau+T} h_1(t, s) x(s) ds + \iint_{(\tau-m)}^{(\tau+T)} h_2(t, s_1, s_2) x(s_1) x(s_2) ds_1 ds_2, \\ \tau \leq t \leq \tau + T, \quad \tau \in R^1, \tag{4.25}$$

$$z(t) = y(t) + v(t), \tag{4.26}$$

where the kernels h_1 and h_2 are square-integrable in all their arguments over any bounded set; where x is square integrable on any finite interval; where $m \geq 0, T > 0$ are arbitrary but fixed, and where $v(t)$ is a wide-sense stationary stochastic process with mean zero. The pair (h_1, h_2) represent the system, x is the input, v is noise, and z is the observed output. T is the duration of an observation interval and m is the maximum duration of the finite memory.

Let τ be temporarily set equal to zero; let $\{\phi_i\}$ and $\{\eta_i\}$ be complete orthonormal systems (c.o.n.s.) for $L^2(-m, T)$ and $L^2(0, T)$, respectively. Expansion of both sides of (4.25) with respect to these c.o.n.s. yields

$$c_k = \sum_i b_{ki} a_i + \sum_{i_1, i_2} b_{ki_1 i_2} a_{i_1} a_{i_2}, \quad k = 1, 2, \dots, \tag{4.27}$$

where

$$a_n = (x, \phi_n), \\ c_n = (y, \eta_n),$$

$$b_{ki} = \int_0^T \int_{-m}^T h_1(t, s) \phi_i(s) \eta_k(t) ds dt, \\ b_{ki_1 i_2} = \int_0^T \int_{-m}^T \int_{-m}^T h_2(t, s_1, s_2) \phi_{i_1}(s_1) \phi_{i_2}(s_2) \eta_k(t) ds_1 ds_2 dt.$$

Since it may be assumed without loss of generality that h_2 is symmetric in s_1 and s_2 , it may be assumed that $b_{k_2 i_2} = b_{k_2 i_1}$. Under rather obvious compactness conditions on the class of systems and class of inputs in question (e.g., if h_1 , h_2 , and x are required to belong to compact subsets of the appropriate L^2 spaces) Eqs. (4.27) can be approximated uniformly in an L^2 sense by a finite system of equations in finitely many unknowns:

$$c_k = \sum_{i=1}^M b_{ki} a_i + \sum_{i_1, i_2=1}^M b_{k i_1 i_2} a_{i_1} a_{i_2}, \quad k = 1, \dots, K. \quad (4.28)$$

These equations can be written in the form of the linear vector-matrix equation $y = Xh$, where h is the vector of the b 's, y is the vector of the c 's, and X is determined by the a 's.

Equations (4.28) correspond to one measurement, as the term is defined above, at $\tau = t_1 = 0$. Successive measurements can be made independently at $t_2 = m + T$, $t_3 = 2(m + T)$, etc. With the c.o.n.s. translated appropriately, each measurement will correspond to a system of equations (4.28) with the input for that measurement determining the a 's and the condition of the system during that measurement determining the b 's. If the b 's do not change for a sufficient number of measurements, input signals can always be chosen so that the b 's can be uniquely determined (see Fiske, 1975b).

Now we suppose that the system in question is causal, and we temporarily also assume it to be time-invariant. The ϕ_i and η_i can then be chosen to take advantage of these conditions. In particular, define ϕ_1, \dots, ϕ_M by

$$\begin{aligned} \phi_i(t) &= \left(\frac{M}{T+m} \right)^{1/2}, & -m + \frac{(i-1)(T+m)}{M} \leq t < -m + \frac{i(T+m)}{M}, \\ &= 0, & \text{for all other } t \in [-m, T]. \end{aligned}$$

The completion of the system $\{\phi_i\}$ is irrelevant. There is no harm done if m is increased if necessary so that it is exactly divisible by $\delta = (T+m)/M$. The η_i can now be defined to be functions consisting of a simple step of length δ , just as are the ϕ_i . However, because of the causality, bounded memory and time-invariance there is nothing lost by stopping the observation interval at δ and defining only η_1 as a step of height $(1/\delta)^{1/2}$ on $[0, \delta]$.

Thus, $T = \delta$, $K = 1$, $M = (m/\delta) + 1$, and Eqs. (4.28) reduce to the single equation

$$c = \sum_{i=1}^M b_i a_i + \sum_{i,j=1}^M b_{ij} a_i a_j, \quad (4.29)$$

where

$$c = \frac{1}{\delta^{1/2}} \int_0^\delta y(t) dt,$$

$$b_i = \frac{1}{\delta} \int_{-m+(i-1)\delta}^{-m+i\delta} \int_0^\delta h(t, s) dt ds,$$

$$b_{ij} = \frac{1}{\delta^{3/2}} \int_{-m+(i-1)\delta}^{-m+i\delta} \int_{-m+(j-1)\delta}^{-m+j\delta} \int_0^\delta h(t, s_1, s_2) dt ds_1 ds_2.$$

It will be observed that, since $b_{ij} = b_{ji}$, there are $N = \frac{1}{2}M(M + 3)$ unknown b 's in (4.29). We therefore consider a block of N successive measurements with N different inputs (i.e., N different sets of the a_i) chosen so that the b 's are uniquely determined. Using a superscript to denote the number of the measurement, we then have from (4.29)

$$c^n = \sum_{i=1}^M b_i a_i^n + \sum_{i,j=1}^M b_{ij} a_i^n a_j^n, \quad n = 1, \dots, N. \tag{4.30}$$

Equations (4.30) can be rewritten in vector-matrix form,

$$y = Xh, \tag{4.31}$$

where

$$y = [c^1, c^2, \dots, c^N]^T,$$

$$h = [b_1, b_{11}; b_2, b_{22}; \dots; b_M, b_{MM}; 2b_{12}, 2b_{13}, \dots, 2b_{1M}; 2b_{23}, \dots; 2b_{M-1, M}]^T,$$

and

$$X = \begin{bmatrix} a_1^1, (a_1^1)^2, a_2^1, (a_2^1)^2, \dots, (a_M^1)^2, a_1^1 a_2^1, \dots, a_{M-1}^1 a_M^1 \\ \dots \\ a_1^N, (a_1^N)^2, \dots, a_1^N a_2^N, \dots, a_{M-1}^N a_M^N \end{bmatrix}.$$

From (4.26) and (4.31), there follows the equation

$$z = Xh + v, \tag{4.32}$$

where $v = [v_1, \dots, v_N]^T$, and v_i is equal to $\delta^{-1/2}$ times the integral of the noise over the observation interval of the i th measurement. $E v_i = 0$, $E(v_i)^2 = \text{constant} = \sigma^2$, say; and we shall assume $E(v_i v_j) = 0$ for $i \neq j$. Thus, $E(vv^T) = \sigma^2 I$.

If it is now presumed that h (and hence the b 's) changes somewhat from one block of N measurements to the next, and if the same cycle of inputs

is used for each block of measurements, we recover Eqs. (4.1), where the superscript n denotes the number of the block of measurements. Theorem 2 applies with $R = \sigma^2 I$ and

$$C = (X^T R^{-1} X)^{-1} X^T R^{-1} = X^{-1}.$$

As an illustration, consider the simplest nontrivial case with $M = 2$. Then $N = 5$ so that five measurements to a block are necessary. The inputs may be chosen so that

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then,

$$C = X^{-1} = \frac{1}{4} \begin{bmatrix} 2 & -2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ -4 & 0 & -4 & 0 & 4 \end{bmatrix}.$$

The eigenvalues and corresponding eigenvectors of $(X^T R^{-1} X)^{-1} = \sigma^2 (X^T X)^{-1}$ are

$$\begin{aligned} \sigma_1^2 &= 3.33\sigma^2, & \psi_1 &= [0.166, 0.166, 0.166, 0.166, -0.944]^T, \\ \sigma_2^2 &= \frac{1}{2}\sigma^2, & \psi_2 &= (1/2^{1/2})[-1, 1, 0, 0, 0]^T, \\ \sigma_3^2 &= \frac{1}{2}\sigma^2, & \psi_3 &= (1/2^{1/2})[0, 0, -1, 1, 0]^T, \\ \sigma_4^2 &= \frac{1}{2}\sigma^2, & \psi_4 &= \frac{1}{2}[1, 1, -1, -1, 0]^T, \\ \sigma_5^2 &= 0.15\sigma^2, & \psi_5 &= [0.474, 0.474, 0.474, 0.474, 0.331]^T. \end{aligned}$$

The η_i remain to be chosen. If limits on the possible variation of $(b_1, b_{11}, b_2, b_{22}, 2b_{12})$ between blocks are taken to be $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$, then the η_i should be

$$\eta_i = \max_{|\alpha_j| \leq \theta_j} [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]^T [\psi_i].$$

In any given situation the choice of the η_i will probably be something of a guess. If they are chosen conservatively (too large), then the estimates will not be quite as good as they might be; if they are chosen too small the error bounds will be incorrect. Suppose in this example the ratio $\eta_i/\sigma =$

0.5. Then $\eta_i/\sigma_1 = 0.28$, and from Fig. 1 one sees that the upper bound on the final mean-squared error in the ψ_1 -component is approximately $0.5\sigma_1^2 = 1.7\sigma^2$, as contrasted with $\sigma_1^2 = 3.33\sigma^2$, the error given by the LUMV estimator.

Lower Bound on Maximum Estimation Error

As previously mentioned, since the error bound just derived is presumably not the best possible, it is important to establish a lower bound on upper bounds to mean-squared error, for comparison. This is done by a contradiction argument using standard results from the Kalman linear least-mean-squared error estimation theory. The result is for linear estimators only.

Suppose that all the conditions for Theorem 2 are satisfied, and in addition each w^n is a stochastic vector satisfying the conditions

$$\begin{aligned} P[(w^n, \psi_i) = \eta_i] &= 1/2 = P[(w^n, \psi_i) = -\eta_i], \\ E[(w^n, \psi_i)(w^m, \psi_j)] &= \eta_i^2 \delta_{ij} \delta_{nm}, \\ E[(w^n, f)(w^m, g)] &= 0, \end{aligned} \quad (4.33)$$

for $i, j = 1, \dots, p$, all m and n , and all f and $g \in R^p$ in the last equation. Note that these additional conditions in no way violate the original ones. Then $Ew^n = 0$ and w^n has covariance Q satisfying

$$(Q\psi_i, \psi_j) = \eta_i^2 \delta_{ij}. \quad (4.34)$$

The equations (4.1) can now be interpreted as representing a controllable, observable linear dynamical system model with uncorrelated state and observation noise. To apply the linear LMS recursive estimation theory to (4.1) it is necessary to assume that h^1 is a stochastic vector with known mean and covariance. It is convenient to take

$$\begin{aligned} Eh^1 &= Cz^1 = \hat{h}^1, \\ \text{cov}(h^1 - Eh^1) &\stackrel{a}{=} P_1 = (X^T R^{-1} X)^{-1}. \end{aligned} \quad (4.35)$$

The error covariance matrix for the LMS estimate \hat{h}^n of h^n given the observations z^2, \dots, z^n and initial data as in (4.35) satisfies the recurrence equation

$$P_n = (P_{n-1} + Q)\{I - X^T[X(P_{n-1} + Q)X^T + R]^{-1} \cdot X(P_{n-1} + Q)\}, \quad n = 2, 3, \dots \quad (4.36)$$

(See, e.g., Åström, 1970, with P_1 given by (4.35).)

LEMMA 5. *The sequence $\{\text{Tr } P_n\}$, $n = 1, 2, \dots$, with P_n given by (4.35) and (4.36) is monotonically nonincreasing.*

Proof. We note first that $\text{Tr } P_2 \leq \text{Tr } P_1$. In fact, given the data $\{z^2; \hat{h}^1, P_1\}$, $\text{Tr } P_2$ is the mean-squared error of the LMS estimate \hat{h}^2 . But the LUMV estimate of \hat{h}^2 has mean-squared error $\text{Tr}[(X^T R^{-1} X)^{-1}] = \text{Tr } P_1$. Since this is another linear estimate based on (some of) the same data it cannot have smaller mean-squared error than the LMS estimate. Hence $\text{Tr } P_2 \leq \text{Tr } P_1$.

Now observe that because of the stationarity of the statistics in the model, the LMS estimate of \hat{h}^n given $\{z^2, \dots, z^n; \hat{h}^1, P_1\}$ has the same error covariance matrix, P_n , as does the LMS estimate of \hat{h}^{n+1} given $\{z^3, \dots, z^{n+1}; \hat{h}^2 = C z^2, P_1\}$. Thus, given $\{z^2, \dots, z^{n+1}; \hat{h}^1, P_1\}$, one can form a (nonoptimum) linear estimate of \hat{h}^{n+1} by first computing the LUMV estimate $\hat{h}^2 = C z^2$, which has error covariance matrix P_1 , and then replacing \hat{h}^2 in the recursive LMS algorithm by \hat{h}^2 . This is equivalent to finding the LMS estimate of \hat{h}^{n+1} given $\{z^3, \dots, z^{n+1}; \hat{h}^2, P_1\}$. By the remark just made, the mean-squared error of this estimate is $\text{Tr } P_n$, which must then be greater than or equal to the minimum mean-squared error $\text{Tr } P_{n+1}$. ■

Let $e = \lim_{n \rightarrow \infty} \text{Tr } P_{n+1}$, then e is a lower bound on uniform upper bounds on mean-squared error. More precisely, we have:

THEOREM 4. *Let the conditions of Theorem 2 be satisfied. Then for any linear estimator of the system parameters h^n and for any $m \geq 1$ there exists an admissible sequence $\{h^n\}$, $n = 1, 2, \dots, m$ such that the mean-squared error in the estimation of the h^m is greater than or equal to e .*

Proof. We are free to suppose that h^1 is a stochastic vector with covariance P_1 and that the w^n are stochastic vectors satisfying (4.33) and bounded as required. Denote an arbitrary linear estimate of h^n by \hat{h}^n . Suppose that for some $m \geq 1$

$$E\{\|\hat{h}^m - h^m\|^2 \mid h^1, w^1, \dots, w^{m-1}\} < e,$$

for every h^1 and every sequence $\{w^1, \dots, w^{m-1}\}$ meeting the conditions just imposed. Then

$$E\|\hat{h}^m - h^m\|^2 = E[E\{\|\hat{h}^m - h^m\|^2 \mid h^1, w^1, \dots, w^{m-1}\}].$$

But by Lemma 5 the LMS estimates of the h^n for the given conditions all have mean-squared error $\text{Tr}(P_n) \geq e$. Thus, by the minimum mean-squared error property of the LMS estimates there is a contradiction. ■

For the very special case that X is the identity this lower upper bound can easily be computed. We compute it here because it gives an immediate

comparison with the upper bound of Theorem 2, and also because, simple as it is, it is the case of interest in the general theorem of the last section.

With $X = I$ the ψ_i , $i = 1, \dots, p$, are orthonormal eigenvectors of R . Thus, if they are chosen as a basis for a vector-matrix formulation, R is the matrix $\text{diag}[\sigma_1^2, \dots, \sigma_p^2]$, and from (4.34) Q is the matrix $\text{diag}[\eta_1^2, \dots, \eta_p^2]$. Furthermore, (4.36) then becomes, after a small amount of manipulation,

$$P_n = (P_{n-1} + Q)(P_{n-1} + Q + R)^{-1}R, \quad n = 2, 3, \dots \quad (4.37)$$

Since $P_1 = R$ is diagonal, each P_n as given by (4.37) is diagonal. In particular, if $P_{n-1} = \text{diag}[\beta_{(n-1)1}^2, \beta_{(n-1)2}^2, \dots, \beta_{(n-1)p}^2]$, then $P_n = \text{diag}[\beta_{n1}^2, \beta_{n2}^2, \dots, \beta_{np}^2]$, where

$$\beta_{ni}^2 = \frac{(\beta_{(n-1)i}^2 + \eta_i^2) \sigma_i^2}{\beta_{(n-1)i}^2 + \eta_i^2 + \sigma_i^2}, \quad i = 1, \dots, p. \quad (4.38)$$

It follows from Lemma 4 that the β_{ni}^2 converge monotonically to the fixed point of the function

$$f(x) = \frac{(x + \eta_i^2) \sigma_i^2}{x + \eta_i^2 + \sigma_i^2}, \quad x > 0,$$

which is

$$\beta_i^2 = -\frac{1}{2}\eta_i^2 + \frac{1}{2}\eta_i^2 \left(1 + 4\frac{\sigma_i^2}{\eta_i^2}\right)^{1/2}. \quad (4.39)$$

Thus,

$$e = \lim_{n \rightarrow \infty} \text{Tr}(P_n) = \sum_{i=1}^p \beta_i^2. \quad (4.40)$$

See Fig. 1 for a comparison with the upper error bound for the recursive estimator.

5. IDENTIFICATION OF A SLOWLY VARYING SYSTEM WITH NOISY OBSERVATIONS OF OUTPUT

As previously indicated, there is really no way to distinguish between h^n and ϵ^n in (3.11) in estimates based on z^n . So, in analogy to what was done in the noise-free case, we estimate $h^n + \epsilon^n$ and simply allow for the error possibly introduced by ϵ^n when the estimate obtained is used as an estimate of h^n . From (3.7) $y^n = h^n + \epsilon^n$, so (3.11) can be rewritten in the form

$$\begin{aligned} y^{n+1} &= y^n + \gamma^n, \\ z^n &= y^n + v^n. \end{aligned} \quad (5.1)$$

Since $\|\gamma^n\| = \|(y^{n+1} - h^{n+1}) + (h^{n+1} - h^n) + (h^n - y^n)\|$ it follows from Lemma 3 that if $\|w_n\| \leq \eta$ then

$$\|\gamma^n\| \leq 2(N - 1)\eta + N\eta = d, \tag{5.2}$$

where d is defined by (5.2). We define an estimate for \hat{h}^n of h^n in (3.11) to be the estimate \hat{y}^n of y^n in (5.1) specified as in Theorem 2. The following theorem is an extension of Theorem 1 that accounts for the statistical errors in the system parameter estimates when the output observations are noisy.

THEOREM 5. *Let $(\mathcal{Y}_e, H^e, \mathcal{X}_e)$ be an extended system satisfying all the conditions of Theorem 1. Suppose that the output for each measurement is observed in the presence of additive noise V_k , as in (3.9), and that the Hilbert-space-valued random variables V_k satisfy the conditions (α) , (β) , (γ) following (3.9). Fix $\epsilon > 0$ arbitrarily and let N be as in Theorem 1. Then, given any $\epsilon' > 0$, there exists an estimate \hat{H}_k of H_k , for any k , formed from a finite sequence of blocks of N measurements such that*

$$\|\hat{H}_n - H_n\| \leq \epsilon + (N - 1)N^{1/2}\delta + \zeta, \tag{5.3}$$

where ζ is a nonnegative random variable satisfying

$$E\zeta^2 \leq \sum_{i=1}^{MN} B_i^2 + \epsilon'.$$

The length of the sequence of blocks depends on ϵ' . Each constant B_i is the unique nonnegative solution of

$$\frac{(B_i^{1/2} + d)^2 \sigma_i^2}{(B_i^{1/2} + d)^2 + \sigma_i^2} = B_i, \quad i = 1, \dots, MN, \tag{5.4}$$

where the σ_i^2 , $i = 1, \dots, MN$, are the eigenvectors of $\pi\Gamma\pi$ (each σ_i^2 is repeated N times), and where

$$d = 2(N - 1)N^{1/2}\delta + N^{3/2}\delta. \tag{5.5}$$

Proof. We maintain the convention that superscripts refer to blocks of measurements. Put $\hat{H}^n = \psi_1(\hat{h}^n) = \psi_1(\hat{y}^n)$. From Theorem 1 it follows that if the estimate of \hat{y}^n is perfect, $\hat{H}^n = \bar{H}^n$, where \bar{H}^n satisfies (3.8). From Lemma 2 it follows that if $\|\hat{h}^n - y^n\| \leq \zeta$, then

$$\begin{aligned} \|\hat{H}^n - H^n\| &\leq \|\hat{H}^n - \bar{H}^n\| + \|\bar{H}^n - H^n\| \\ &\leq \zeta + \epsilon + (N - 1)N^{1/2}\delta, \end{aligned}$$

since $\psi_1(\hat{h}^n - y^n) = \hat{H}^n - \bar{H}^n$. This much is true no matter what estimate is used for h^n ; however, consider \hat{h}^n as defined above. From Theorem 2

$$E \|\hat{h}^n - h^n\|^2 \leq \sum_{i=1}^p b_{ni}^2,$$

where now $p = MN$, and where the b_{ni} are defined recursively by (4.21) and (4.22). Since $X = I$, the σ_i^2 in these equations are now simply the eigenvalues of the covariance operator $R = \pi\Gamma\pi$. Since (5.2) is to hold, each η_i , $i = 1, \dots, p$, is replaced by d , which, in terms of δ is as in (5.5). The assertion then follows from Theorem 3 for H^n , which corresponds to the nN th measurement. The conclusion also follows for any H_n ; one can simply delay the beginning of the first block of measurements by the appropriate amount so that $n = kN$ for some k . ■

RECEIVED: May 30, 1975; REVISED: January 19, 1976

REFERENCES

- ALBERT, A. (1972), "Regression and the Moore-Penrose Pseudoinverse," pp. 86-93, Academic Press, New York.
- ÅSTRÖM, K. J. (1970), "Introduction to Stochastic Control Theory," pp. 225-237, Academic Press, New York.
- BARUCHA-REID, A. T. (1972), "Random Integral Equations," pp. 14-16, Academic Press, New York.
- DIEUDONNÉ, J. A. (1960), "Foundations of Modern Analysis," pp. 126-129, Academic Press, New York.
- FISKE, P. H., AND ROOT, W. L. (1974), Estimation of parameters in systems of relatively slow time variation, in "Proceedings of the Twelfth Annual Allerton Conference on Circuit and System Theory," pp. 1039-1047, Univ. of Illinois Press, Urbana, Illinois.
- FISKE, P. H. (1975a), Modeling and identification of time-varying systems from noisy observations, in "Proceedings of the Johns-Hopkins Conference on Information Science and Systems," The Johns Hopkins University, Baltimore.
- FISKE, P. H. (1975b), "Modeling and Identification of Relatively Slowly Varying Systems," Ph. D. Thesis, The University of Michigan.
- ROOT, W. L. (1971), Some general structure theory of systems to be used in identification and measurement, in "Proceedings of the Fifth Annual Princeton Conference on Information Sciences and Systems, pp. 13-19, Princeton University, Princeton, N. J.
- ROOT, W. L. (1973), On the modelling and estimation of communication channels, in "Multivariate Analysis, III" (Krishnaiah, Ed.), pp. 61-78, Academic Press, New York.

- Root, W. L. (1975a), On the modeling of systems for identification, I: ϵ -Representations of classes of systems, *SIAM J. Control* 13, 927-944.
- Root, W. L. (1975b), On the modeling of systems for identification, II: Time-varying systems, *SIAM J. Control* 13, 945-974.
- SCHWEPPE, F. C. (1973), "Uncertain Dynamic Systems," pp. 154-160, Prentice-Hall, Englewood Cliffs, N. J.