

## On the Structure of a Class of System Identification Problems\*

Sur la structure d'une catégorie de problèmes d'identification des systèmes

Über die Struktur einer Klasse von Problemen der Systemidentifikation

О структуре класса проблем опознавания систем

W. L. ROOT†

*The identification of non-linear systems from input and output data is essentially a linear problem, so that general questions of when and how can a system be identified can be answered with linear analysis.*

**Summary**—Systems of the form  $y=Hx$  are considered, where  $x$  is the input signal,  $y$  is the output and  $H$  is a transformation. The input and output signals belong to  $L_2$  spaces,  $H$  belongs to a specified class of continuous, not necessarily linear, transformations. The fundamental problem is to identify  $H$  from measurements of  $x$  and  $y$ .

The approximate representation of certain transformations  $H$  in terms of polynomial integral operators is discussed. The concept of a determinable class, which is used throughout, and some of its properties are reviewed. An illustrative example concerning human operator tracking problems is presented. A structural characterization of determinable classes based on work of Hsieh and Balakrishnan is given and then used as foundation for an application of linear regression analysis to the problem of identifying an  $H$  belonging to a determinable class when there is observation noise. Finally, the notion of determinable class is applied to a class of stochastic operators and an elementary convergence theorem is given.

### 1. INTRODUCTION

THE SUBJECT of this paper is the identification of systems as represented in input–output form. The purpose is to present some theoretical structure to serve as a foundation for the analysis of a rather wide class of identification problems, with special emphasis on the questions, when, and in what sense is identification possible. The systems considered may be in general linear or non-linear, time-invariant or time-varying, and with single or multi-variable inputs and outputs, although some restrictions are imposed in places. Stochastic systems are also considered. The work here is intended to apply

chiefly to off-line testing of processes, or to fixed experiments on systems under controlled conditions. No effort is made to treat iterative identification procedures which up-date and improve the information about an unknown system as it operates. However, some of the material of this paper is relevant to a study of such procedures.

The background of the work presented here is contained chiefly in earlier work by PROSSER and the author [1, 2, 4, 5] and by BALAKRISHNAN and HSIEH [3], for example. Specific connections are pointed out later.

Section 2 contains general definitions and notations, and the idea is introduced that an input–output identification problem can always be treated as a linear problem. This observation is fundamental for all that follows. In section 3 some basic facts are given about polynomial integral transformations, such as a Volterra series, and an approximation theorem is stated. Such transformations are used throughout the rest of the paper to represent the systems being considered, and the material of section 3 is included partly to justify their use. See also a comment in the Conclusions section. Section 4 is illustrative example couched in terms of a human operator tracking problem. Section 5 contains a solution for the noise-free identification problem in certain cases in terms of the spectral resolution of an operator determined by the input signal. This is essentially an abstract version of a portion of a development done by Hsieh and Balakrishnan, and is included because sections 6 and 7 depend on it. In section 6 a special modification of infinite-dimensional linear regression analysis is applied to identification when there is observation noise present. Section 7 contains definitions and a theorem about determining the moments of an unknown stochastic system. Both sections 6 and

\* Received 8 April 1970; revised 9 October 1970. The original version of this paper was presented at the IFAC Symposium on Identification and Process Parameter Estimation which was held in Prague, Czechoslovakia during June 1970. It was recommended for publication in revised form by associate editor B. Gaines.

This work was partly supported by the Office of Naval Research Grant Nr. N00014-67-A-0181-0021.

† The University of Michigan, Ann Arbor, Michigan, U.S.A. The author was visiting the University of Cambridge, Cambridge, United Kingdom when this paper was prepared.

7 are believed to be formally new, although the mathematical arguments required are rather obvious.

These last three sections actually present algorithms for identification, but they are in terms of eigenvalue problems which would normally be too difficult to solve. They are really intended to provide existence theorems and a foundation for the study of efficient computational algorithms in special cases. The procedure described in the example, section 4, on the other hand, is practical in many circumstances, but it is fairly standard. The concept of a determinable class of unknown systems, defined in section 2, is used throughout to give a unified point of view.

## 2. DEFINITIONS AND PRELIMINARY RESULTS

The systems to be considered can be characterized by the equation  $y=H(x)$ , where  $x$  (the input signal) is an element of a specified set  $\mathcal{X}$ ,  $y$  (the resulting output signal) is an element of a specified set  $\mathcal{Y}$ , and  $H$  (the system transformation) is an element of a specified class  $\mathcal{H}$  of functions from  $\mathcal{X}$  into  $\mathcal{Y}$ . The basic identification problem to be discussed is, briefly stated, to find an  $H \in \mathcal{H}$  that adequately represents an unknown system from the results of one or a series of experiments involving the introduction of known input signals  $x$  and the measurement of the corresponding output signals  $y$ . The sets  $\mathcal{X}$ ,  $\mathcal{Y}$ ,  $\mathcal{H}$ , together with any mathematical structure they are to have, will be stipulated as the paper proceeds. In section 7 the transformations  $H$  are stochastic.

One can regard each  $x \in \mathcal{X}$  as determining a function from  $\mathcal{H}$  into  $\mathcal{Y}$ . To emphasize this point of view, we write  $y=H(x)=X(H)$ ,  $H \in \mathcal{H}$ , i.e.  $X$  is the mapping that assigns output  $y$  to system  $H$  when  $x(\sim X)$  is the input signal and  $y=H(x)$ . Then, obviously, the problem of identifying  $H$  when the input signal is  $x$  is essentially the problem of inverting the mapping  $X$ . If the output space  $\mathcal{Y}$  is a linear space this can always be interpreted as a *linear* inversion problem as follows. Define addition and scalar multiplication in the way they are conventionally defined for spaces of functions, i.e. for each  $H_1, H_2, H \in \mathcal{H}$  and all real numbers  $\alpha$  define  $H_1 + H_2, \alpha H$ , respectively, by

$$(H_1 + H_2)(x) = H_1(x) + H_2(x), \quad x \in \mathcal{X}$$

$$(\alpha H)(x) = \alpha[H(x)], \quad x \in \mathcal{X}.$$

Extend  $\mathcal{H}$  if necessary so that it is closed under linear combinations; then  $\mathcal{H}$  becomes a linear space, or a vector space. This extension is permissible because it only enlarges the basic class of systems being considered. The notation  $\mathcal{H}$  will be retained, but it is henceforth assumed  $\mathcal{H}$  is a

full linear space. Then  $X$  is a linear transformation on  $\mathcal{H}$ , because:

$$X(\alpha H_1 + \beta H_2) = (\alpha H_1 + \beta H_2)(x) = \alpha[H_1(x)]$$

$$+ \beta[H_2(x)] = \alpha X(H_1) + \beta X(H_2).$$

The first and third equalities follow from the definition of  $X$ , the second from the definitions of addition and scalar multiplication in  $\mathcal{H}$ . In summary:

(A) With the linear structure prescribed for  $\mathcal{H}$ , the mapping  $X$  from  $\mathcal{H}$  into  $\mathcal{Y}$  defined by  $y = X(H) = H(x)$  is linear.

This simple observation, which is obvious to anyone who has thought about such things, is important because it says that the input-output identification problem can always be studied by linear analysis. When observation noise is added, the problem can be treated as a linear regression problem. The observation takes on practical as well as theoretical importance if, when parameters are introduced, as they must be to permit actual computation, they are introduced in such fashion as to determine linearly the transformations  $H$ . Then the actual computation algorithm will be linear, or, in the usual terminology, the problem will be "linear in the parameters". The only condition necessary for this linearity of  $X$  is that  $\mathcal{Y}$  be a linear space, and this condition is satisfied when  $\mathcal{Y}$  is an  $L_2$ -space as required in the next paragraph.

For the remainder of the paper it will be assumed that the input and output signals  $x(t)$  and  $y(t)$  will be real- or vector-valued functions that are square-integrable Lebesgue on  $R^1$ , the real line, or some subset thereof. For  $x(t)$  to be vector-valued will mean that  $x(t) = \{x_1(t), \dots, x_M(t)\}$ ,  $M$  finite, where each  $x_i(t)$  is real-valued. In case  $x(t), z(t)$  are vector-valued square-integrable functions on a set  $A$  the notations

$$|x(t)|^2 = x_1^2(t) + \dots + x_M^2(t),$$

$$\|x\|^2 = \int_A |x(t)|^2 dt$$

and

$$(x, z) = \int_A \sum_{i=1}^M x_i(t)z_i(t) dt$$

are used. We shall denote the  $L_2$ -space with the inner product just defined by  $L_2(A)$  for both the scalar-valued and vector-valued cases. It will be indicated in context when signals may be vector-valued. The set  $\mathcal{H}$  will always be a subset of the set of continuous functions from  $\mathcal{X}$  into  $\mathcal{Y}$ , where the topologies of  $\mathcal{X}$  and  $\mathcal{Y}$  are those given by the

$L_2$  norms. Two different norms will be used for  $\mathcal{H}$ : the supremum norm defined by

$$\|H\| = \sup_x \|Hx\|,$$

useful when  $\mathcal{X}$  is bounded, and a Hilbert space norm, to be defined later, which can be used when  $\mathcal{H}$  contains only transformations of a certain form.

Although the kind of identification problem in question here is basically the inversion of the transformation  $X$ , there are several general considerations which influence the directions to be taken. First, the transformations  $X$  are usually not invertible; so we must be content with something like pseudo inverses, or we must limit their domains so that they are (nearly) invertible. Second, even if the inversion is possible in principle, a precise determination of  $H$  is usually neither necessary nor practically feasible. So even before taking into account statistical errors caused by noisy observations, there is some point in regarding the identification problem as an approximation problem. Third if the identification algorithm is to be carried out by digital computation, it must involve only finite sets of data. Even if the algorithm is to be done by an analogue technique, there is often some insight provided by considering finite-dimensional approximations. In an attempt to construct some abstract identification theory that explicitly takes account of these considerations, the author and others [1, 2, 4, 5-7] have used the concept of a *determinable class of unknown systems*. There are of course other ways to do this, but the determinable class concept will be used here. A formal definition follows. However, it should be noted that in the view of the author this definition should be regarded as a prototype, to be modified, generalized or specialized in obvious ways as may be appropriate to a given situation.

Let  $\mathcal{X}$  be a subset of an  $L_2$ -space; let  $\mathcal{Y}$  be an  $L_2$ -space, and let  $\mathcal{H}$  be a normed space of continuous functions  $H$  with domain  $\mathcal{X}$  and taking values in  $\mathcal{Y}$ . The norm of  $H$ , which is arbitrary as far as this definition is concerned, is written  $\|H\|$ . Let  $\mathcal{D} \subset \mathcal{H}$ , then:

**Definition 1.** Let  $\varepsilon > 0$  be given.  $\mathcal{D}$  is an  $\varepsilon$ -*determinable class* if there exists  $x \in \mathcal{X}$ , a finite set of continuous real-valued functionals  $m_i(y)$ ,  $i = 1, \dots, N$ , bounded on bounded sets, and a continuous function  $\hat{H}$  from  $R^N$  ( $N$ -dimensional real Euclidean space) into  $\mathcal{H}$  such that  $\|H - \hat{H}(m_1(Hx), \dots, m_N(Hx))\| \leq \varepsilon$  for all  $H \in \mathcal{D}$ .

**Definition 1a.**  $\mathcal{D}$  is a *determinable class* if it is  $\varepsilon$ -determinable for every  $\varepsilon > 0$  ( $x$ ,  $N$ ,  $\{m_i\}$  and  $\hat{H}$  may all depend on  $\varepsilon$ ).  $\mathcal{D}$  is *determinable with respect to* (wrt)  $x$  if it is  $\varepsilon$ -determinable for every

$\varepsilon > 0$  with the same  $x$ . The value taken on by  $\hat{H}$  is a *determination* of the unknown system.

The idea behind these definitions is to formalize criteria as to when systems can be identified from input-output measurements, because in general they cannot, and to establish a point of view from which to consider identification procedures. The definitions do not encompass the question of random errors of observation, which must be considered separately, and, since they are for deterministic systems only, they will hence have to be modified when we consider stochastic systems.

It is easy to show [1] that a necessary condition for a bounded set  $\mathcal{D}$  to be determinable is that it be totally bounded\* in the metric topology given by the norm, or, equivalently, that  $\mathcal{D}$  have compact closure. This is by no means sufficient, however, for it does not guarantee the approximate invertibility the definition requires. A necessary and sufficient condition, proved in Ref. [1] is:

- (B) A bounded subset  $\mathcal{D}$  of  $\mathcal{H}$  is determinable if and only if: (i)  $\mathcal{D}$  is totally bounded (ii) for any  $\varepsilon > 0$  there is an  $x = x(\varepsilon) \in \mathcal{X}$  such that for any  $H_1$  and  $H_2 \in \mathcal{D}$  satisfying  $\|H_1 - H_2\| > \varepsilon$ , then  $\|H_1x - H_2x\| > 0$ .

Superficially, the determinable class definition applies to identification using measurements with only one input signal—we shall call this *one-shot identification*, but this is not really so. Of course, for time-invariant systems it is fundamentally meaningless to distinguish between one-shot and  $K$ -shot identification if the choice of the observation interval  $T$  is free. However, for time-varying systems it is not meaningless, if we consider for the  $K$ -shot identification that the time variation is repeated for each new testing period. Also, for systems with finite memory, even time-invariant ones, it is conceptually useful to consider repeated measurements instead of one continuous measurement. A little thought will convince one that the determinable class concept applies to  $K$ -shot identification also, if one regards the sequences of input and output signals as new signals represented as vectors of length  $K$ .

A remark is in order before concluding this section. The term identification can have various

\* A subset  $\mathcal{D}$  of a metric space is totally bounded if, given any  $\varepsilon > 0$ ,  $\mathcal{D}$  can be covered with a finite number of open spheres of radius  $\varepsilon$  with centers in  $\mathcal{D}$ . In a separable Hilbert space (e.g. an  $L_2$ -space) an equivalent condition is: given any complete orthonormal set  $\{\phi_i\}$ , and any  $\varepsilon > 0$ ,  $\mathcal{D}$  is totally bounded if and only if i)  $\|x\| \leq \text{constant}$  for all  $x \in \mathcal{D}$ , ii) there is a positive integer  $N = N(\varepsilon)$  such that

$$\sum_{i=N}^{\infty} |(x, \phi_i)|^2 \leq \varepsilon$$

for all  $x \in \mathcal{D}$ . In a finite-dimensional Euclidean space any bounded set is totally bounded.

connotations, including that of finding the actual physical structure of an unknown system. Here, obviously, we do not ask for so much; we ask only for some mathematical representation of a transformation, close in some norm to the transformation performed by the system. For arbitrary norms this may be a little artificial, but if the norm is the sup norm it is evidently reasonable—because then we know that  $\|H - H\| < \epsilon$  is exactly the right condition to guarantee that having made the identification  $H$  we can predict the output of the system to within  $\epsilon$  for any input  $x \in \mathcal{X}$  as long as the system remains unchanged.

### 3. POLYNOMIAL INTEGRAL TRANSFORMATIONS

In the remaining sections the systems being discussed are represented by transformations of the form:

$$\begin{aligned} y(t) &= [H(x)](t) \\ &= \sum_{n=0}^N \int_A \dots \int_A h_n(t, s_1, \dots, s_n) x(s_1) \\ &\quad \dots x(s_n) ds_1 \dots ds_n \quad t \in T \end{aligned} \quad (1)$$

where  $A$ , the time during which the input is applied and  $T$ , the time during which the output is observed, are subsets of  $R^1$ , where  $x \in L_2(A)$ , where the zero-th order term is  $h_0(t)$ , and where  $N$  is any finite positive integer. In the case of vector-valued signals,  $y(t) = \{y_1(t), \dots, y_M(t)\}$ , and the expression  $h_n(t, s_1, \dots, s_n) x(s_1) \dots x(s_n)$  is to be interpreted as the vector

$$\begin{aligned} \sum_{i_1, \dots, i_n}^M h_n^{i_1, j_1, \dots, i_n}(t, s_1, \dots, s_n) x_{i_1}(s_1) \dots x_{i_n}(s_n) \\ j = 1, \dots, M'. \end{aligned}$$

Note that each  $h_n$  may be taken to be symmetric in the variables  $s_1, \dots, s_n$  since only the symmetric part of  $h_n$  contributes to the integral. One can justify the generality of such transformations on intuitive grounds, since the class of all such transformations is obviously a large class, by considering infinite series and analytic functionals, or by considering approximations to continuous functions using the Stone-Weierstrass theorem. Here the last is done, and although the basic idea is well-known there are some details to be considered, so a formal statement is made. First some elementary properties are established.

Let the  $n^{\text{th}}$  term of the sum in equation (1) be denoted  $H_n(x)$ .  $H_n$  is a homogeneous polynomial integral transformation, and

$$H = \sum_{n=1}^N H_n.$$

Define a norm for  $H_n$ , to be denoted by  $\|H_n\|$ , by

$$\begin{aligned} \|H_n\|^2 &= \sum_{j=1}^{M'} \sum_{i_1, \dots, i_n=1}^M \int_T \int_A \dots \int_A \\ &\quad |h_n^{i_1, j_1, \dots, i_n}(t, s_1, \dots, s_n)|^2 ds_1 \dots ds_n dt. \end{aligned}$$

Then define  $\|H\|$  by

$$\|H\|^2 = \sum_{n=1}^N \|H_n\|^2.$$

The norms so defined will be called Hilbert-Schmidt norms ( $\|H_n\|$  is the conventional Hilbert-Schmidt norm in the linear case when  $n=1$ ). Both norms are obviously true norms, the first being the ordinary norm for the  $L_2$ -space of vector-valued functions of  $(n+1)$  variables,  $\mathcal{L}_n = L_2(A^n \times T)$ , and the second being the norm of the direct sum of the  $\mathcal{L}_n$ ,  $\mathcal{L} = \mathcal{L}_0 \oplus \dots \oplus \mathcal{L}_N$ . It is consistent to use the same notation for both norms, since the norm for the homogeneous transformations can be regarded as a special case of the other one. For transformations as given by equation (1) with finite Hilbert-Schmidt norm the following properties hold:

(C) If  $y = H_n(x)$ ,  $\eta = H_n(\xi)$ ,  $x$  and  $\xi \in L_2(A)$ , and  $\|H_n\| < \infty$  then:

$$(i) \quad y \in L_2(T), \text{ and } \|y\| \leq \|H_n\| \|x\|^n.$$

$$(ii) \quad \|y - \eta\| \leq \sqrt{2} n \|H_n\| \|x - \xi\| [\max(\|x\|, \|\xi\|)]^{n-1}.$$

*Proof.* These follow from the Schwarz inequality. (i) is trivial. See Ref. [2], e.g. for a proof of (ii) in the scalar case.

It follows from (C) that for any fixed  $N$  and with each  $\|H_n\| < \infty$ ,  $H$  as given by equation (1) is a uniformly continuous function from any bounded subset  $\mathcal{X}$  of  $L_2(A)$  into  $L_2(T)$ . Let  $\mathcal{X}$  be any fixed compact subset of  $L_2(A)$  and let  $L_2(T)$  be fixed, then denote by  $\mathcal{H}_N$  the class of all polynomials of the form of equation (1) of degree less than or equal to  $N$  mapping  $\mathcal{X}$  into  $L_2(T)$ . Let  $\mathcal{F}$  be the class of all bounded continuous functions  $F$  from  $\mathcal{X}$  into  $L_2(T)$  made into a Banach space with the usual norm:

$$\|F\| = \sup_{x \in \mathcal{X}} \|Fx\|.$$

Then  $\mathcal{H}_N \subset \mathcal{F}$ , and the following approximation theorem holds

(D) Let  $\mathcal{C} \subset \mathcal{F}$  have compact closure. Then, given  $\epsilon > 0$ , there exists  $N = N(\epsilon)$  such that for every  $F \in \mathcal{C}$ , there is an  $H \in \mathcal{H}_N$  satisfying  $\|F - H\| \leq \epsilon$ .

*Proof.* See Ref. [8].

This approximation theorem is taken to be sufficient justification for restricting the discussion

to determinable classes of systems where the transformations  $H$  are all of the form given by equation (1), even though the signal space,  $\mathcal{X}$  is not always assumed to be compact.

4. EXAMPLE

To help clarify the concepts already introduced and to provide a lead-in to the material of the next three sections, an illustrative example involving a second-degree polynomial integral transformation is discussed. The calculations to be made actually apply for the most part to any system representable by such a transformation, however the example will be formulated in terms of compensatory tracking problems for a human operator of a dynamical system.

In a compensatory tracking situation, a human operator adjusts the magnitude and polarity of a signal forcing a given dynamical system so as to try to make the system output follow a given input. Let the output signal be denoted by  $y(t)$  and the input signal by  $x(t)$ ; both are real-valued functions of time. The operator does not know  $x(t)$ , but is presented with a real-time visual display of the error,  $\epsilon(t) = x(t) - y(t)$ . He operates a manual device, say a stick, that controls the signal into the dynamical system and tries to reduce the error to zero. Suppose for simplicity that the signal into the dynamical system is proportional to the position of the stick, and let it be denoted by  $z(t)$ . The human operator is acting as the controller in a closed-loop control system as shown in Fig. 1. In this environment the human operator is an unknown system with input  $\epsilon(t)$  and output  $z(t)$ . One problem is to identify this unknown system which takes  $\epsilon(t)$  into  $y(t)$ ; another is to identify the complete closed-loop system which takes  $x(t)$  into  $y(t)$ . For convenience, these are referred to respectively as problems I and II. As will be indicated, problem I may not be very practical, but it does illustrate certain points in the paper.

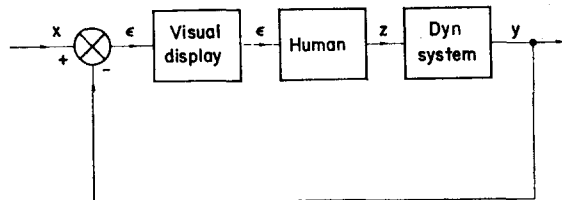


FIG. 1. Human operator tracking loop.

In the usual approach to problem I the human operator is characterized as a time-invariant linear dynamical system of known form, but with a few unknown parameters to be determined. It is known, however, that neither of the assumptions, time-invariance or linearity, is quite true, and even

if they were the assumption of the form of the transfer characteristic for the human operator is pretty much guesswork. So it would be reasonable to consider a more general model for the human operator, and a model that allows for time variation and some nonlinearity is:

$$z(t) = \int_0^\delta g_1(t, s)\epsilon(s)ds + \int_0^\delta \int_0^\delta g_2(t, s_1, s_2)\epsilon(s_1)\epsilon(s_2)ds_1ds_2, \quad 0 \leq t \leq \tau. \tag{2}$$

It is not known to the author from experimental results whether in fact this is or is not an adequate and reasonable model for the human operator. However, it is plausible, and permits bringing out certain points in discussion.

If the controlled system is linear or nearly so, a plausible model for problem II is, similarly,

$$y(t) = \int_0^\delta h_1(t, s)x(s)ds + \int_0^\delta \int_0^\delta h_2(t, s_1, s_2)x(s_1)x(s_2)ds_1ds_2, \quad 0 \leq t \leq \tau. \tag{3}$$

Now there is a real difference between the two problems in that the input signal,  $x(t)$ , can be predetermined by the experimenter in problem II, whereas  $\epsilon(t)$ , the input for problem I, cannot be. Some consequence of this fact are mentioned below. However, the form of equations (2) and (3) are the same, so the formal calculations to follow apply to either.

Let the following assumptions be made:

- (i)  $0 < \delta \leq \tau < \infty$ . Note that in problem I  $\delta = \tau$ , in problem II the input signal can be turned off and the operator given some additional time so that  $\tau$  can be strictly greater than  $\delta$ .
- (ii) with a suitable normalization,

$$\|x\|^2 = \int_0^\delta x^2(t)dt \leq 1.$$

Note that this can be guaranteed in problem II and made reasonably certain in I.

- (iii)  $|h_1(t, s)|$  and  $|h_2(t, s_1, s_2)|$  are each bounded by fixed constants, which is a reasonable assumption for either problem. This condition implies that  $\|H^2\| = \|h_1\|^2 + \|h_2\|^2$ , as previously defined, is finite, since the ranges of integration are finite.
- (iv)  $h_1(t, s) = 0$  for  $t < s$ ,  $h_2(t, s_1, s_2) = 0$  for  $t < s_1$  or  $t < s_2$ , i.e. the system is causal.

Now let  $\{\phi_i(s)\}$ ,  $i=1, 2, \dots$ , be a complete orthonormal system (cons) for  $L_2[0, \delta]$  and  $\{\eta_i(t)\}$ ,  $i=1, 2, \dots$ , be a cons for  $L_2[0, \tau]$ . A specific choice of these functions will be made later. Then there exist the generalized Fourier expansions:

$$x(s) = \sum_{n=1}^{\infty} a_n \phi_n(s), \quad a_n = (x, \phi_n) \quad (4)$$

$$y(t) = \sum_{n=1}^{\infty} c_n \eta_n(t), \quad c_n = (y, \eta_n) \quad (5)$$

$$h_1(t, s) = \sum_{j, i=1}^{\infty} b_{ji} \eta_j(t) \phi_i(s), \quad (6)$$

$$b_{ji} = \int_0^{\tau} \int_0^{\delta} h_1(t, s) \eta_j(t) \phi_i(s) ds dt$$

$$h_2(t, s_1, s_2) = \sum_{j, i_1, i_2=1}^{\infty} b_{j i_1 i_2} \eta_j(t) \phi_{i_1}(s_1) \phi_{i_2}(s_2), \quad (7)$$

$$b_{j i_1 i_2} = \int_0^{\tau} \int_0^{\delta} \int_0^{\delta} h_2(t, s_1, s_2) \eta_j(t) \phi_{i_1}(s_1) \phi_{i_2}(s_2) ds_1 ds_2 dt.$$

It should be observed that there is of course a slight misuse of notation here; in each case the partial sums in the expansions on the right converge in mean square to the left hand side, e.g.

$$\lim_{N \rightarrow \infty} \left\| \sum_{n=1}^N a_n \phi_n - x \right\|^2 \rightarrow 0;$$

they do not necessarily converge at each  $s$  and  $t$  unless some further condition is imposed.

Using the expansions given by equations (4), (5), (6) and (7) in equation (3) yields

$$y(t) = \sum_{j, i=1}^{\infty} b_{ji} a_i \eta_j(t) + \sum_{j, i_1, i_2=1}^{\infty} b_{j i_1 i_2} a_{i_1} a_{i_2} \eta_j(t),$$

and then forming the inner products  $(y, \eta_k)$  gives the infinite set of equations

$$c_k = \sum_{i=1}^{\infty} b_{ki} a_i + \sum_{i_1, i_2=1}^{\infty} b_{k i_1 i_2} a_{i_1} a_{i_2}, \quad k=1, 2, \dots \quad (8)$$

It will be observed that this is a system of linear equations in the unknown parameters  $b_{ki}$  and  $b_{k, i_1, i_2}$ , as has to be the case since the system representation is linear in  $h_1$  and  $h_2$ , and  $h_1$  and  $h_2$  are linearly dependent on their generalized Fourier coefficients, the  $b$ 's. It is obvious that this property of linearity in the unknown parameters will always hold when a representation of the unknown system in polynomial integral transformations is used.

Now let it be assumed that the unknown system belongs to a determinable, and hence totally bounded, class  $\mathcal{D}$ . Then, by the characterization of

totally bounded sets in a Hilbert space given before, for any  $\varepsilon > 0$  there are  $N = N(\varepsilon)$  and  $N' = N'(\varepsilon)$  such that  $H$  is changed by less than  $\varepsilon$ , in the Hilbert-Schmidt norm, if the series expansions for  $h_1$  and  $h_2$  given by equations (6) and (7) are truncated by allowing  $j$  to run only from 1 to  $N'$  and each  $i$  to run only from 1 to  $N$ . With this truncation the infinite system of equations in infinitely many unknowns, equation (8) is replaced by the finite system of equations:

$$c_k = \sum_{i=1}^N b_{ki} a_i + \sum_{i_1, i_2=1}^N b_{k i_1 i_2} a_{i_1} a_{i_2} + \varepsilon_k, \quad (9)$$

$$k=1, \dots, N'$$

where  $\varepsilon_k$  is an error. From the inequality (i) given in (C) it follows  $|\varepsilon_k| \leq \varepsilon$  for each  $k$ , for all the systems in the class and for any input  $x$  satisfying  $\|x\| \leq 1$ . It is obvious, but important to note, that if the input signal  $x$  can be chosen as desired then it may be chosen so that all the  $a_i$  vanish for  $i > N$  and then the errors  $\varepsilon_k$  in equation (9) all vanish. It should also be noted that  $N(\varepsilon)$  and  $N'(\varepsilon)$ , although always finite, will depend on the choice of the  $\phi_i$ 's and  $\eta_i$ 's, which amounts to saying that the totally bounded class of unknown systems is better described in some coordinate systems than in others.

Thus we are led to consider the system

$$c_k = \sum_{i=1}^N b_{ki} a_i + \sum_{i_1, i_2=1}^N b_{i_1 i_2} a_{i_1} a_{i_2}, \quad (10)$$

$$k=1, \dots, N'$$

which may be exact or approximate, depending on  $x$ . It is desired to find the  $b_{ki}$  and  $b_{k i_1 i_2}$ , for once they are known  $H$  is known to within  $\varepsilon$  in Hilbert-Schmidt norm. Now, because of the symmetry of the right hand sides of equations (10) in the  $a_i$ , the  $b_{k i_1 i_2}$  enter only through the symmetric sums  $(b_{k i_1 i_2} + b_{k i_2 i_1})$  and there is no loss in generality in assuming  $b_{k i_1 i_2} = b_{k i_2 i_1}$  for all  $k, i_1, i_2$ . Note that this is the same as assuming  $h_2(t, s_1, s_2)$  is symmetric in  $s_1$  and  $s_2$ . Hence, there are actually  $N' \{N + [N(N+1)/2]\} = (N'/2)(N^2 + 3N)$  unknown  $b$ 's in the  $N'$  equations. Thus, if the class  $\mathcal{D}$  is to be one-shot determinable there must be prior information sufficient to give  $(N'/2)(N^2 + 3N) - N'$  independent relations among the  $b$ 's. At the other extreme of no prior information, the  $b$ 's clearly cannot be determined in fewer than  $(N^2 + 3N)/2$  separate experiments. In fact, one can easily show that  $(N^2 + 3N)/2$  different inputs  $x$  can be chosen so that if that many experiments are performed independently the resulting  $N'[(N^2 + 3N)/2]$  equations are uniquely solvable for the  $b$ 's. To keep things simple, only an example of a choice of inputs for

the case  $N=3$  is given, but the result generalizes to all  $N$  in a completely obvious way. Actually only blocks of  $(N^2+3N)/2$  equations have to be considered simultaneously when there are no extra constraints, because the equations for one fixed value of  $k$  are independent of those for another.

In the case  $N=3$ , for each  $k$  there are nine equations in nine unknowns. Let the  $n^{\text{th}}$  input be

$$x^{(n)}(t) = a_1^{(n)}\phi_1(t) + a_2^{(n)}\phi_2(t) + a_3^{(n)}\phi_3(t).$$

Assign values to the  $a$ 's according to the following table:

	$a_1$	$a_2$	$a_3$
1	1	0	0
2	2	0	0
3	0	1	0
4	0	2	0
5	0	0	1
6	0	0	2
7	1	1	0
8	1	0	1
9	0	1	1

If the matrix of coefficients of the  $b$ 's, for each fixed  $k$ , is now written out it is readily seen to be nonsingular. The equations decouple so that at most two equations need be solved simultaneously, and the solutions are:

$$b_{k1} = \frac{4c_k^{(1)} - c_k^{(2)}}{2}, \quad b_{k11} = \frac{c_k^{(2)} - 2c_k^{(1)}}{2}$$

$$b_{k2} = \frac{4c_k^{(3)} - c_k^{(4)}}{2}, \quad b_{k22} = \frac{c_k^{(4)} - 2c_k^{(3)}}{2}$$

$$b_{k3} = \frac{4c_k^{(5)} - c_k^{(6)}}{2}, \quad b_{k33} = \frac{c_k^{(6)} - 2c_k^{(5)}}{2}$$

$$b_{k12} = c_k^{(n)}, \quad b_{k13} = c_k^{(8)}, \quad b_{k33} = c_k^{(9)}$$

where  $c_k^{(n)} = (y^{(n)}, \eta_k)$  and  $y^{(n)}$  is the output corresponding to the input  $x^{(n)}$ . It is clear that although if there are many unknowns many experiments are required, the computations can be very simple if the input signals are properly chosen.

Everything that has been said thus far is true no matter what the orthonormal functions  $\phi_i$  and  $\eta_i$  may be. However, three particular choices seem worth mentioning: ordinary Fourier series of sines and cosines; Haar functions, or functions closely related); and eigen-functions of particular operators, to be discussed in general in the next Section. Haar functions, it will be recalled, are defined as plus-minus couplets of square waves on successive intervals of length  $2^{-n}$ ,  $n=1, 2, \dots$

However, the Haar functions do not actually need to appear here because the first  $2^n$  Haar functions on  $[0, \tau]$  for any positive integer  $n$ , can be replaced by the functions

$$\phi_i(t) = 0, \quad 0 < t < \frac{(i-1)\tau}{2^n}$$

$$= 2^{n/2}, \quad \frac{(i-1)\tau}{2^n} < t < \frac{i\tau}{2^n}$$

$$= 0, \quad \frac{i\tau}{2^n} < t < \tau, \quad i=1, 2, \dots, 2^n$$

in the sense that the  $\phi_i$  are orthonormal and can be made into a cons by adjoining the Haar functions from  $2^{n+1}$  on. In context,  $N$  is taken to be  $2^n$  for some  $n$ , and the  $\phi_i$  defined as above. With these  $\phi_i$  the  $a$ 's become simply averaged sample values of  $x$ , which are easy to compute and interpret and if the  $\psi_i$  are chosen similarly, the  $b$ 's and  $c$ 's also become averaged sample values. These functions are particularly useful if the unknown system is time invariant or even nearly so, for then there are simple relations among the  $b$ 's which reduce the number of unknowns. In the pure time-invariant case, using these  $\phi_i$  and  $\eta_i$ , the  $b_{ki}$  can be written as  $b_{k-i}$ , and the  $b_{ki_1i_2}$  as  $b_{k-i_1, k-i_2}$ .

Thus far the discussion has been concerned entirely with the noise-free case, and in fact observation noise will not be treated in connection with this example. However, if there is noise added to the observation only, i.e. if  $y$  is replaced by  $y+n$  where  $n$  is noise, then the estimation of the  $b$ 's becomes a standard linear regression problem to be treated in the usual way. The fact that the system of equations (10) is invertible for a proper choice of inputs guarantees that each  $b$  can be estimated, and this is discussed further in section 6. If the input is perturbed by additive noise then the estimation is no longer a linear regression, of course; this problem is not considered here. In the human operator identification problems additive noise might well not be a serious factor; however, if additive noise must be considered anywhere in the closed loop then it affects the input, and the linear regression analysis based on equation (10) does not apply. Practically, problem I is more likely to be involved in difficulty, because under normal operation the input  $\varepsilon(t)$  is small so that small noise disturbances are important. This is just a manifestation of the fact that with any identification scheme at all the identification of an element interior to a closed loop is chancy at best.

It is apparent that the human operator identification problem formulated as problem II can be solved in the noise-free case by the procedure indicated, but that repeated measurements are

usually necessary, that is, there will be  $k$ -shot determinability. The number of measurements depends on how many terms are necessary to describe the unknown system to within the allotted tolerance  $\varepsilon$ , and on how much prior information there is, to be used to relate certain of the  $b$ 's to each other. If the orthogonal functions based on the Haar system are used, one can readily specify from a rough knowledge of the response time of a human operator what length of sampling interval is necessary, and hence how many orthogonal functions and how many parameters are necessary. For problem I the procedure may not be very useful, even in the noise-free case, because it is not possible to predetermine the input signals  $\varepsilon(t)$ . This implies two things: the equations (10) become approximations because of the neglected coefficients  $a_i, i > N$ , which has the same sort of effect as noise at the input, and the  $a_i$  will not be such as to permit an economical inversion procedure for the system of equations (10). In fact, extra measurements may be required in order to make the system invertible.

There remains a difficulty, which indeed was one of the reasons for choosing this example for illustration, that perhaps a human operator is not nearly enough a repeatable system, so that two experiments run with identical inputs may not yield closely the same output. If this proves to be so, then the human operator may be regarded as a stochastic system as discussed in section 7. His average response may be found by averaging the results of several experiments with the same input, and correlations may be found by the procedures of section 7.

##### 5. DETERMINABLE CLASSES IN TERMS OF SPECTRAL DECOMPOSITIONS

For a rather wide subclass of transformations of the form of equation (1) it is possible to introduce coordinate functions so as to "diagonalize" the identification problem. These functions, which are the eigenfunctions of operators related to  $X$ , replace the  $\eta_i$ 's and the products of  $\phi_i$ 's that appeared in the example just discussed.

From now on it is assumed that  $\mathcal{H}$ , the class of systems to be admitted into discussion, consists of transformations  $H$  of the form of equation (1). Under the change of variables  $\{s_i, t\} \rightarrow \{u_i, t\}$ ,  $u_i = t - s_i$ , and the definition  $k_n(t, u_1, \dots, u_n) = h_n(t, t - u_1, \dots, t - u_n)$ , equation (1) becomes  $y(t)$

$$= \sum_{n=0}^N \int_B \dots \int_B k_n(t, u_1, \dots, u_n) x(t - u_1) \dots x(t - u_n) du_1 \dots du_n, \quad t \in T \quad (11)$$

where the  $u_i$  range through  $B$ . We now make some new special assumptions: (1) the zero'th order

term is zero,  $k_0 = 0$ ; (2) either the kernels are time-invariant,  $k_n(t, u_1, \dots, u_n) = k_n(u_1, \dots, u_n)$ , or  $k_n$  is of the special form

$$k_n = \sum_{i=1}^M \alpha_{n,i}(t) k_{n,i}$$

where each  $k_{n,i}$  is time-invariant and the  $\alpha_{n,i}(t)$  are known bounded functions; (3)  $T$  is a finite interval in  $R^1$ . The reason for (1) is that in almost any problem a zero'th degree term can either be found in advance and subtracted out, or else it is random and represents observation noise and so should be treated specially. Noise is considered in the next section. The reason for (2) is that it simplifies the problem and permits the development of the kind of structure theory given here. It does not quite mean we are limited to time-invariant systems, since time-varying systems with "finitely many degrees of freedom" in the time variation are allowed for by the generalization indicated. See also the last section. The reason for (3) is that observation periods always are finite. This assumption also happens to imply a discrete spectral theory, as will be seen. It will be assumed here that  $k_n(u_1, \dots, u_n) = 0$  if any  $|u_i| > c = \text{finite constant}$ . Such systems will be referred to as systems with finite memory, although of course such a restriction implies finite anticipation also. In such cases the set  $B$  can certainly be taken to be  $[-c, c]$ . As before, each  $x \in L_2(A)$ , and each  $k_n$ , or  $k_{n,i}$  in the time-varying generalization,  $\in L_2(B^n)$ . Then it follows from (C) that  $y \in L_2(T)$ .

A system transformation  $H$  is now represented as the family of time-invariant kernels  $k = k\{k_1, \dots, k_N\}$ , or, in the time-varying case as  $k = \{k_{11}, \dots, k_{1M_1}, \dots, k_{N1}, \dots, k_{NM_N}\}$ . In the first instance the transformation  $X$  can be written as a sum of linear integral transformations

$$y = Hx = Hk = X_1 k_1 + \dots + X_N k_N \quad (12)$$

where  $X$  is a linear mapping from the space  $\mathcal{H} = \mathcal{L}_1 \oplus \dots \oplus \mathcal{L}_N$  into  $L_2(T)$ , where now and henceforth  $\mathcal{L}_n = L_2(B^n)$ , and where  $X$  and each  $X_n$  are bounded linear transformations. In fact, under the assumption that  $T$  is a finite interval:

(E)  $X$  is a compact transformation.

*Proof.* Since  $x \in L_2(A)$ , each  $X_n$  is Hilbert-Schmidt, by an obvious calculation, as long as  $T$  has finite measure, irrespective of whether  $A$  has finite measure or not. Hence each  $X_n$  is compact, from which it follows easily that  $X$  is compact.

In the second instance  $Xk$  can again be written as a sum, of which  $a_{ni}(t) X_{ni} k_{ni}$  is a typical term. Since  $a_{ni}(t)$  is bounded and  $X_{ni}$  is compact, the product  $a_{ni}(t) X_{ni}$  is compact, and  $X$  is compact.



Thus (E) still holds. For convenience, when expressions need to be written out we shall consider only the time-invariant case; the changes necessary to go to the more general case will be clear.

It follows immediately from (E) by standard Hilbert space operator theory, for example, see Ref. [9], that:

(F) (i)  $G = X^*X$  is a compact, self-adjoint, positive operator from  $\mathcal{H}$  into  $\mathcal{H}$ . Hence there exists an orthonormal set of eigenvectors  $\phi_i \in \mathcal{H}$  and corresponding eigenvalues  $\lambda_i > 0$ ,  $\lambda_i \rightarrow 0$ , such that  $G\phi_i = \lambda_i\phi_i$  and such that for any  $h \in \mathcal{H}$   $Gh = \sum c_i\phi_i$ . (ii)  $F = XX^*$  is a compact, self-adjoint, positive operator on  $L_2(T)$ . If  $\eta_i = \lambda_i^{-1/2}X\phi_i$ , then  $\{\eta_i\}$  is an orthonormal set in  $L_2(T)$ ,  $F\eta_i = \lambda_i\eta_i$ , and for any  $y \in L_2(T)$ ,  $Fy = \sum d_i\eta_i$ . (iii) Let  $\mathcal{S}_x$  be the closed linear subspace of  $\mathcal{H}$  spanned by the  $\phi_i$ . Then  $\overline{R(X^*)}$  (the closure of the range of  $X^*$ ) =  $\mathcal{S}_x$ . The null space of  $G$  = the null space of  $X = \mathcal{S}_x^\perp$  (the orthogonal complement of  $\mathcal{S}_x$ ).

Thus the problem of inverting the operator  $X$  to find  $k$  has the following, necessarily incomplete, solution:

(G) Let  $k = k_x + k'$ ,  $k_x \in \mathcal{S}_x$ ,  $k' \in \mathcal{S}_x^\perp$ . Let  $k_x = \sum a_i\phi_i$ . Then  $y = Xk = \sum a_i\lambda_i^{1/2}\eta_i$ . Since the  $\lambda_i$  are known, in principle, the  $a_i$  can be found from  $(\eta_i, y) = a_i\lambda_i^{1/2}$ , and  $k_x$  is then known.  $k'$  cannot be found from operations on  $y$ .

This solution was essentially obtained by HSIEH and BALAKRISHNAN, e.g. Ref. [3]. From (G) one has (see Ref. [4] for a proof, which is however almost immediate) a characterization of classes determinable wrt  $x$ :

(H) Let  $\mathcal{D} \subset \mathcal{H}$  be bounded and determinable wrt  $x$ , where the norm on  $\mathcal{H}$  used for the definition of determinability is defined by

$$\|H\|^2 = \sum_{n=1}^N \|k_n\|^2,$$

where  $\|k_n\|$  is the regular norm for  $L_2(B^n)$ . Then, for  $\mathcal{C}$  some totally bounded subset of  $\mathcal{S}_x$ ,  $\mathcal{D} = \{k_x + f(k_x)\}$  where  $k_x \in \mathcal{C}$  and  $f(k_x)$  is a known continuous function from  $\mathcal{C}$  into  $\mathcal{S}_x^\perp$ . Conversely, any subset of this form is a class determinable wrt  $x$ .

It follows from the definition that any determinable class can be approximated by a class determinable wrt  $x$ , and hence a set of the form just described. If  $\mathcal{X}$  is bounded so that the norm

$$\|H\| = \sup_{x \in \mathcal{X}} \|Hx\|,$$

is appropriate, then it also follows from the definition that if is determinable in  $\|\cdot\|$  norm, the Hilbert-Schmidt norm, it is also determinable in  $\|\cdot\|$  norm (sup norm).

To save a considerable amount of space and also to make the arguments more evident, the development in this Section has been given in an abstract operator notation. The operators  $X^*X$ , written out, are in fact  $N \times N$  matrix arrays of polynomial integral operators and the  $\phi_i$  are  $N$ -vectors of functions, the first of one variable, the second of two variables, etc. For example, see Refs. [3, 4] for the full expressions.

### 6. OBSERVATION NOISE

With observation noise present the identification problem of the previous section becomes that of estimating  $k$  in the equation

$$y(t) = (Xk)(t) + n(t), \quad t \in T \tag{13}$$

where  $n(t)$ , which represents the noise, is either a sample function from a real-valued (measurable, separable) stochastic process, or, in the vector case, is the set of sample functions

$$n(t) = \{n_1(t), \dots, n_M(t)\}.$$

We assume  $En_i(t) = 0$ ,  $t \in T$ , and  $R_{ij}(t, \tau) = En_i(t)n_j(\tau)$ ,  $i, j = 1, \dots, M$ , is continuous in the closure of  $T \times T$ . Thus all sample functions except a set of probability zero belong to  $L_2(T)$  since  $T$  is of finite measure. The set of functions  $k = \{k_1, \dots, k_N\}$  are to satisfy the same conditions as in section 5.

A modification of linear regression analysis is given that is fitted to the determinable class concept. As will be seen, the condition that  $H$ , or  $k$ , belongs to a determinable class greatly simplifies the noise analysis, since it reduces the problem to an essentially finite-dimensional one.

For both theoretical and computational purposes it is convenient to recast equation (13) so that the linear operation is self-adjoint. Thus we may look at either

$$y = XX^*b + n, \quad k = X^*b \tag{14}$$

or at

$$g = X^*y = X^*Xk + X^*n \tag{15}$$

Although the first has one advantage of being stated in a simpler space the second leads to a simpler exposition. We note that  $X^*n$  is a "noise" in  $\mathcal{H} = \mathcal{L}_1 \oplus \dots \oplus \mathcal{L}_N$  whose matrix covariance function defines a non-negative definite, self-adjoint, compact operator  $\Gamma$  on  $\mathcal{H}$  satisfying  $E(X^*n, f)(X^*n, g) = (\Gamma f, g)$ ,  $f, g \in \mathcal{H}$ .

Let  $\varepsilon > 0$  be chosen arbitrarily small, but then fixed. Since  $k$  is stipulated to belong to a determinable class  $\mathcal{D}$ ,  $k_x$ , the projection of  $k$  on  $\mathcal{S}_x$  must belong to a fixed totally bounded subset of  $\mathcal{S}_x$ . There is then an  $N = N(\varepsilon)$  such that if

$k_x = k'_x + k''_x$ , where  $k'_x$  is the projection of  $k_x$  on the finite-dimensional linear subspace  $\mathcal{S}'_x$  spanned by  $\{\phi_1, \dots, \phi_N\}$ ,  $\|k_x - k'_x\| < \varepsilon$  for all  $k \in \mathcal{D}$ . We content ourselves with estimating  $k'_x$ , since if  $k'_x$  were truly known we would have an  $\varepsilon$ -determination of  $k_x$ , the only part of  $k$  that can be found without supplementary information. Now define the operator  $S$  on  $\mathcal{S}'_x$  as the restriction of  $X^*X$  to  $\mathcal{S}'_x$ . Obviously, if  $k_x = k'_x + k''_x$ ,  $k'_x \in \mathcal{S}'_x$ ,  $k''_x \in \mathcal{S}_x \ominus \mathcal{S}'_x$ ,  $X^*Xk_x = Sk'_x + r$ , where  $Sk'_x \in \mathcal{S}'_x$  and  $r \in \mathcal{S}_x \ominus \mathcal{S}'_x$ . Let  $\{\psi_1, \dots, \psi_N\}$  be an orthonormal set of eigenvectors of  $S^{-1}\Gamma S^{-1}$ , which is a self-adjoint operator on  $\mathcal{S}'_x$ , and let  $\{\mu_1, \dots, \mu_N\}$  be the corresponding eigenvalues,  $\mu_i > 0$ ,  $i = 1, \dots, N$ . Let  $P'_x$  be the projection on  $\mathcal{S}'_x$  and consider the restriction of equation (15),

$$g' = P'_x g = Sk'_x + P'_x X^* n. \quad (16)$$

Operating on equation (9) with  $S^{-1}$  gives

$$f = S^{-1}g' = k'_x + S^{-1}P'_x X^* n. \quad (17)$$

Then,

$$f = \sum (k_x, \psi_i) \psi_i + \sum \xi_i \psi_i$$

where the  $\xi_i$  are the random variables

$$(S^{-1}P'_x X^* n, \psi_i).$$

The  $\xi_i$  satisfy:

$$\begin{aligned} E\xi_i &= 0, \\ E\xi_i \xi_j &= E(X^* n, S^{-1}\psi_i)(X^* n, S^{-1}\psi_j) \\ &= (S^{-1}\Gamma S^{-1}\psi_i, \psi_j) = \mu_i \delta_{ij}. \end{aligned}$$

The only estimator linear in  $f$  that gives an unbiased estimate of  $k'_x$  is

$$\hat{k} = \sum (f, \psi_i) \psi_i.$$

This estimate gives a mean squared error of

$$E\|k_x - \hat{k}\|^2 = \sum \mu_k. \quad (18)$$

However, depending on  $\mathcal{D}$ ,  $x$  and the noise, there may be obvious linear estimators that give uniformly a smaller mean squared error. In particular it may be better to truncate the linear unbiased estimator  $\hat{k}$  by deleting terms. Let  $\alpha$  be a subset of the integers  $1, \dots, N$  and  $\beta$  be the complementary subset. Define

$$\hat{k}_{\alpha} = \sum_{i \in \alpha} (f, \psi_i) \psi_i.$$

Then

$$E\|\hat{k} - k'_x\|^2 = \sum_{i \in \alpha} \mu_k + \sum_{j \in \beta} |(k'_x, \psi_j)|^2.$$

Since  $\mathcal{D}_x$  is totally bounded and its projection on  $\mathcal{S}'_x$  is already "thin", it may be possible to ensure that certain  $\{(k'_x, \psi_j)\}^2$ , or sums of such terms, are less than the corresponding  $\mu_j$ , or sums of  $\mu_j$ 's. In such case  $\beta$  should include these  $j$ 's. To summarise:

(1) Given  $\varepsilon > 0$ , there exists a linear estimator  $\hat{k}$  of  $k_x$  with bias  $\leq \varepsilon$  and mean-squared error

$$E\|\hat{k} - k_x\|^2 < \sum \mu_k + \varepsilon^2,$$

where  $N = N(\varepsilon)$  and  $\mu_k$  are as defined above. In certain cases one can exhibit linear estimators with uniformly less mean-squared error over the determinable class.

The (formal) special case when the observation noise is white is valid and yields the result

$$E\|k'_x - \hat{k}\|^2 = \sum (\sigma_o^2 / \lambda_k),$$

where  $\sigma_o^2 = E|(n, \eta_i)|^2$  for all  $i$ .

## 7. STOCHASTIC SYSTEMS STATIONARY IN THE OBSERVATION TIME

This section consists of a slight elaboration on the observation that if the kernels of equations of the form of equation (11) are stochastic, then taking second moments of these equations yields new equations of the same general form, of twice the degree, with deterministic kernels. The concept of determinable classes can be applied in a statistical sense to such stochastic systems by applying it in the usual way to the related deterministic systems formally described by the second moment equations. The chief reason for suggesting this approach is that it provides a framework for studying the identification, in some sense at least, of a class of time-varying systems. In general, the identification of time-varying systems is possible only if the systems are known to have special properties, and the definition that follows prescribes one special property sufficiently restrictive to at least open up the possibility of stochastic identification.

A system describable by equation (11) is said to be *observation-time stationary* (ots) if the fourth moments of all the kernels  $k_n(t, u_1, \dots, u_n)$  exist and are bounded for all values of the variables and if  $E k_n(t + \tau, u_1, \dots, u_n) k_m(t, u'_1, \dots, u'_m) = \Gamma_{nm}(\tau; u_1, \dots, u_n; u'_1, \dots, u'_m)$ ,  $n, m = 0, 1, \dots, N$ , where each  $\Gamma_{nm}$  depends on observation times only through their difference. The  $\Gamma_{nm}$  are bounded for all values of their arguments, *a fortiori*, by the assumption about fourth moments. A practical example of a linear ots system is given by the scattering of electromagnetic waves from a volume containing a great many randomly moving charged particles where such parameters as the number and

velocity distribution of the particles are kept constant. Another possible example is given by human operator problems, as discussed previously.

Consider its stochastic systems representable by equation (11) with real-valued inputs and outputs, with the following additional constraints on the stochastic kernels: (i)  $k_n(t, u_1, \dots, u_n) = 0$  with probability one whenever  $|u_i| > c$ ,  $i = 1, \dots, n$ ;  $n = 1, \dots, N$  (finite memory and anticipation). (ii)  $\Gamma_{00}(\tau)$  is known, and  $\Gamma_{0n}(\tau; u_1, \dots, u_n) = 0$ ,  $n = 1, \dots, N$ . (iii)  $E k_n(t, u_n, \dots, u_n) = 0$ ,  $n = 0, 1, \dots, N$ . Condition (ii) is imposed because  $k_0(t)$  is taken to represent observation noise; it is of no consequence mathematically. Condition (iii) is imposed for convenience; if the average values of the kernels are not zero one would estimate their average values, in a fashion analogous to the estimation of covariances to be discussed, and subtract the estimates. The basic observation interval, denoted by  $T$ , will be taken to be  $[0, T]$ ,  $T > 0$ , but it is necessary to consider repeated experiments with successive observation intervals of length  $T$ . The input signal  $x$  will belong to  $L_2(A)$  where  $A$  is an interval  $[-c-b, T+c+b]$  for some  $b > 0$ .

Now let  $\tau_0$ ,  $0 < \tau_0 < b$ , be fixed and let  $d = \tau_0 + c$ . Then, for any  $\tau$  satisfying  $|\tau| < \tau_0$ .

$$\begin{aligned} y(t+\tau) &= k_0(t+\tau) + \sum_{n=1}^N \int_{-c}^c \dots \int_{-c}^c k_n(t+\tau; u_1, \dots, u_n) x(t+\tau-u_1) \\ &\quad \dots x(t+\tau-u_n) du_1 \dots du_n, \quad t \in T, \\ &= k_0(t+\tau) + \sum_{n=1}^N \int_{-d}^d \dots \int_{-d}^d k_n(t+\tau; v_1+\tau, \dots, v_n+\tau) x(t-v_1) \\ &\quad \dots x(t-v_n) dv_1 \dots dv_n. \end{aligned}$$

Taking the expected value of  $y(t)y(t+\tau)$  yields

$$\begin{aligned} E y(t)y(t+\tau) &= \Gamma_{00}(\tau) + \sum_{p=1}^N \int_{-d}^d \dots \int_{-d}^d \left[ \sum_{n=1}^p \Gamma_{n, p-n}(\tau; v_1+\tau, \dots, v_n+\tau; v_{n+1}, \dots, v_p) \right] \\ &\quad x(t-v_1) \dots x(t-v_p) dv_1 \dots dv_p. \end{aligned} \quad (19)$$

We put  $z_\tau(t) = E y(t)y(t+\tau) - \Gamma_{00}(\tau)$  (recall that  $\Gamma_{00}(\tau)$  is assumed to be known), and

$$\begin{aligned} \gamma_{p, \tau}(v_1, \dots, v_p) \\ = \sum_{n=1}^p \Gamma_{n, p-n}(\tau; v_1+\tau, \dots, v_n+\tau; v_{n+1}, \dots, v_p). \end{aligned}$$

Then equation (19) becomes:

$$\begin{aligned} z_\tau(t) &= \sum_{p=1}^N \int_{-d}^d \dots \int_{-d}^d \gamma_{p, \tau}(v_1, \dots, v_p) x(t-v_1) \\ &\quad \dots x(t-v_p) dv_1 \dots dv_p. \end{aligned} \quad (20)$$

Equation (20) formally describes unknown, time-invariant systems of the kind discussed in section 5, but with the unknown systems represented by the families of kernels  $\gamma_\tau = \{\gamma_{1, \tau}, \dots, \gamma_{N, \tau}\}$ . Of course the interpretation is quite different; since  $z_\tau(t)$  cannot be measured directly but can only be estimated statistically from the results of repeated measurements, the kernels  $\gamma_{p, \tau}$  can at best only be estimated statistically. Note that the conditions imposed are more than sufficient to guarantee that  $\gamma_{\tau \in \mathcal{H}} = \mathcal{L}_1 \oplus \dots \oplus \mathcal{L}_N$ .

We shall say that a class  $\mathcal{D}$  of unknown its stochastic systems each representable by an equation of the form of equation (11) is *second-order stochastically  $\varepsilon$ -determinable* ( $\tau_0$ ) if the fourth moments of the  $k_n(t, u_1, \dots, u_n)$  are uniformly bounded for all  $k \in \mathcal{D}$ , if all  $k_n(t, u_1, \dots, u_n)$  are independent of  $k_m(t', u_1, \dots, u_m)$  for  $|t-t'| > \Delta > T + 2c + 2b$ , and if for each  $\tau, |\tau| < \tau_0$ , the corresponding class of (deterministic) systems formally described by equation (20) is an  $\varepsilon$ -determinable class. Then,

(J) If  $\mathcal{D}$  is such a class of stochastic systems, given  $\varepsilon > 0$ , there exists  $K = K(\varepsilon)$  such that if  $K$  independent measurements are made  $\gamma_\tau, |\tau| < \tau_0$ , can be determined, with probability greater than  $1 - \varepsilon$ , to within  $\varepsilon$  in norm.

*Proof.* Let  $x \in L_2(A)$  give an  $\varepsilon/2$  determination of  $\gamma_\tau$ . With this  $x$ , let  $\rho(t, t+\tau) = E y(t)y(t+\tau)$ . Suppose that measurements are repeated with replicas of  $x$  starting at  $0, \Delta, 2\Delta$ , etc., where  $\Delta > T + 2c + 2b$  so that the outputs  $y$  in any pair of observation intervals are independent. Then

$$\begin{aligned} E \left| \frac{1}{K} \sum_{k=1}^K y(t+k\Delta)y(t+\tau+k\Delta) - \rho(t, t+\tau) \right|^2 \\ = \frac{1}{K} \text{var} |y(t)y(t+\tau)|. \end{aligned} \quad (21)$$

The assumptions of uniformly bounded fourth moments, finite memory and anticipation, and the integrability of  $x^2$  yield the fact, after an obvious calculation, that  $\text{var} |y(t)y(t+\tau)|$  exists, and in fact is uniformly bounded for all  $\tau$  and all  $k \in \mathcal{D}$ . Hence

$$\begin{aligned} \lim_{K \rightarrow \infty} E \int_0^T \left| \frac{1}{K} \sum_{k=1}^K y(t+k\Delta)y(t+\tau+k\Delta) \right. \\ \left. - \rho(t, t+\tau) \right|^2 dt = 0 \end{aligned}$$

uniformly for all  $k \in \mathcal{D}$ , and

$$\lim_{K \rightarrow \infty} E \| \hat{z}_{\tau K} - z_\tau \| = 0$$

uniformly for all  $k \in \mathcal{D}$ , where the estimate  $\hat{z}_{\tau k}$  is given by

$$\hat{z}_{\tau k} = \frac{1}{K} \sum_{k=1}^K y(t+k\Delta)y(t+\tau+k\Delta) - \Gamma_{00}(\tau).$$

Since the mapping from  $\mathcal{Y} \subset L_2(T)$  into  $\mathcal{H}$  that gives the estimate of  $\gamma_\tau$  is uniformly continuous on the image of  $\mathcal{D}$ , by the definition of an  $\varepsilon$ -determinable class, the conclusion follows that for  $K$  sufficiently large  $\hat{z}_{\tau k}$  can be used in a determination of  $\gamma_\tau$  with the stated accuracy.

Evidently, except where the polynomial integral operators are homogeneous, the individual covariances  $\Gamma_{nm}$  of the kernels cannot be estimated; however, this is of no concern since only the  $\gamma_n$  are required to compute output covariances for known inputs. The restriction of attention to covariances, as has been done, is unnecessary, although it would appear to be the only practically useful case; an entirely analogous discussion can be given for the determination of  $n$ th moment functions.

## 8. CONCLUSIONS

A partial theory of the identification of non-linear systems from input-output data has been developed based on: (1) the fact that quite generally the identification problem can be posed as a linear one, (2) the determinable class concept, and (3) representations in terms of polynomial integral transformations, a Volterra series. These ingredients have led to a unified theory covering a reasonably wide class of problems. The actual definition of determinable class is somewhat arbitrary, but some such idea seems necessary to allow for a systematic treatment of approximation errors and reduction of data to finite sets of numbers. The use of Volterra series is subject to some criticism on the practical grounds that it often leads to an unreasonably large number of parameters. However, there are both theoretical and practical reasons for using it. The theoretical reasons are that it is quite general, as pointed out in section 3, that it leads to problem formulations that are linear in the parameters, as evidenced in section 4, and that it is well-fitted to the use of Hilbert Space methods. It is in a sense the natural extension of the linear integral transform formulation of linear problems. It can be practically useful in problems that are only slightly non-linear, as in those mentioned involving a human operator, or in problems that are exactly quadratic, as e.g. many optical filtering problems, or the determination of the covariance of a linear system. However, it is only one representation and others may often be preferable.

The treatment of observation noise made possible by restricting the unknown systems to belong to a determinable class is felt to be of some value,

because it replaces infinite-dimensional linear regression problems, which are badly behaved in the sense that the linear unbiased estimators often have infinite variance, by finite-dimensional ones which, although they do not yield unbiased estimators to the original problem, do yield satisfactory estimators.

Many systems occurring in nature are not repeatable, but they possess a kind of average repeatability which makes it appropriate for them to be modelled as observation-time-stationary systems. The problem of determining covariances of such systems has been embedded in the general identification theory so as to permit the use of results from that theory.

The theory described is only a partial one because among other things, it does not touch on iterative procedures, it does not treat the question of different system representations and it does not concern itself with stochastic inputs.

## REFERENCES

- [1] R. T. PROSSER and W. L. ROOT: Determinable classes of channels. *J. math. Mech.* **16**, 365-398 (1966).
- [2] W. L. ROOT: On system measurement and identification. *Proc. Symp. on System Theory*, Poly. Inst. of Brooklyn, pp. 133-157 (1965).
- [3] H. C. HSIEH: The least squares estimation of linear and nonlinear system weighting function matrices. *Inform. Control* **7**, 84-115 (1964).
- [4] R. T. PROSSER and W. L. ROOT: On the identification of unknown systems. *Proc. Princeton Conf. on Information Sciences and Systems*, pp. 285-289 (1967).
- [5] W. L. ROOT: On the measurement and use of time-varying communication channels. *Inform. Control* **8**, 390-422 (1965).
- [6] P. P. VARAIYA: On determinable classes of signals and linear channels. *SIAM J. appl. Math.* **15**, 440-449 (1967).
- [7] B. S. RAJPUT: Determinable subsets of Banach spaces. Ph.D. Dissertation, Dept. of Math., University of Illinois (1969).
- [8] W. L. ROOT: On the structure of a class of system identification problems. *2nd Prague IFAC Symp. on Identification and Process Parameter Estimation*, Paper 8.1, Prague (1970).
- [9] A. C. ZAAZEN: *Linear Analysis*. North Holland, Amsterdam (1953).

**Résumé**—L'article considère des systèmes de la forme  $y = Hx$ , où  $x$  est la grandeur d'entrée,  $y$  est la grandeur de sortie et  $H$  est une transformation. Les signaux d'entrée et de sortie appartiennent à des espaces  $L_2$ ,  $H$  appartient à une catégorie donnée de transformations continues, non nécessairement linéaires. Le problème fondamental consiste à identifier  $H$  à partir des mesures de  $x$  et de  $y$ .

L'article discute la représentation approchée de certaines transformations  $H$  au moyen d'opérateurs intégraux polynomiaux. Il passe en revue le concept de la catégorie déterminable, qui est constamment utilisé, et quelques unes de ses propriétés. Un exemple se rapportant aux problèmes de l'opérateur humain illustre l'article. Celui-ci donne une caractérisation structurale des catégories déterminables basée sur les travaux de Hsieh et Balakrishnan et cette caractérisation est ensuite utilisée comme base d'application de l'analyse régressive linéaire au problème d'identification d'un  $H$  appartenant à une catégorie déterminable en présence de bruit dans l'observation. Finalement, la notion de catégorie déterminable est appliquée à une catégorie d'opérateurs aléatoires et un théorème élémentaire de convergence est donné.

**Zusammenfassung**—Betrachtet werden Systeme der Form  $y=Hx$ , wobei  $x$  das Eingangssignal,  $y$  das Ausgangssignal und  $H$  eine Transformation ist. Die Eingangs- und Ausgangssignale gehören zu  $L_2$ -Räumen, während  $H$  zu einer vorgeschriebenen Klasse von stetigen nicht notwendigerweise linearen Transformation gehört. Das Grundproblem besteht in der Identifikation von  $H$  durch Messungen von  $x$  und  $y$ . Die angehäufte Darstellung von bestimmten Transformationen in Termen vielgliedriger Integraloperatoren wird diskutiert. Der Begriff einer erfassbaren Klasse, der durchweg benutzt wird, und einige ihrer Eigenschaften werden geprüft. Ein erläuterndes Beispiel in Bezug auf Nachführungsprobleme mit menschlichem Operator wird dargestellt. Eine strukturelle Kennzeichnung erfassbarer Klassen, die auf Arbeiten von Hsiek und Balakrishnan basiert, wird angeführt und als Begründung für eine Anwendung der linearen Regressionsanalyse auf das Problem der Identifikation eines  $H$  angeführt, das zu einer erfassbaren Klasse gehört, wenn ein maßbares Rauschsignal vorhanden ist. Schließlich wird die Vorstellung einer erfassbaren Klasse auf eine Klasse stochastischer Operatoren angewandt und ein elementares Konvergenztheorem angegeben.

**Резюме**—Статья рассматривает системы формы  $y=Hx$ , где  $x$  является входной координатой,  $y$  является выходной координатой и  $H$  является превращением. Входной и выходной сигналы принадлежат к пространствам  $L_2$ ,  $H$  принадлежит к данному классу непрерывных но не обязательно-линейных превращений. Основная задача заключается в опознавании  $H$  и измерении  $x$  и  $y$ .

Статья обсуждает приблизительное изображение некоторых превращений  $H$  при помощи многочленных интегральных операторов. Она рассматривает концепт определяемого класса, который постоянно используется, и некоторые из его свойств. Пример относящийся к проблемам человеческого оператора иллюстрирует статью. Последняя дает структурное охарактеризование определяемых классов основанное на трудах Хсиеха и Балакришнана и это охарактеризование затем используется как основа применения линейного регрессивного анализа к проблеме опознавания некоторого  $H$  принадлежащего к определяемому классу в присутствии шума в наблюдении. В заключение, концепт определяемого класса применяется к классу случайных операторов и дается элементарная теорема сходимости.