ARLINGTON HALL STATION
ARLINGTON 12 VIRGINIA

1 OF 21

...MENT OR OTHER DRAWINGS, SPECIFI... OR ... OTHER DATA
...ISSUED OTHER THAN IN CONNECTION WI... ...CTELY RELATED
...MENT OPERATION, THE U. S. GOVERNM... ...REBY INCURS
...ANY OBLIGATION WHATSOEVER; AND ... ...T THAT THE
...FORMULA... FURNISHED, OR IN ANY ... ...UPPLIED THE
...CATED ...ER DATA IS NOT TO BE ...RIBUTED BY
...WISE ... ...MANNER LICENSING THE ... ...ER OR ANY OTHER
...OR ... ANY RIGHTS OR PRIVILEGES ... TO MANUFACTURE,
...USE ...ATION THAT MAY IN ANY WAY ... ...SLATED THERETO.

THEORY OF THE ANALYSIS OF NONLINEAR SYSTEMS

MARTIN B. BRILLIANT

FC

TECHNICAL REPORT 345

MARCH 3, 1958

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 345 March 3, 1958

THEORY OF THE ANALYSIS OF NONLINEAR SYSTEMS

Martin B. Brilliant

This report is based on a thesis submitted to the Department of Electrical Engineering, M.I.T., January 13, 1958, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

A theory of the analysis of nonlinear systems is developed. The central problem is the mathematical representation of the dependence of the value of the output of such systems on the present and past of the input. It is shown that these systems can be considered as generalized functions, and that many mathematical methods used for the representation of functions of a real variable, particularly tables of values, polynomials, and expansions in series of orthogonal functions, can be used in generalized form for nonlinear systems.

The discussion is restricted to time-invariant systems with bounded inputs. A definition of a continuous system is given, and it is shown that any continuous system can be approximately represented, with the error as small as may be required, by the methods mentioned above. Roughly described, a continuous system is one that is relatively insensitive to small changes in the input, to rapid fluctuations (high frequencies) in the input, and to the remote past of the input.

A system is called an analytic system if it can be exactly represented by a certain formula that is a power-series generalization of the convolution integral. This formula can represent not only continuous systems but also no-memory nonlinear systems. Methods are derived for calculating, in analytic form, the results of inversion, addition, multiplication, cascade combination, and simple feedback connection of analytic systems. The resulting series is proved to be convergent under certain conditions, and bounds are derived for the radius of convergence, the output, and the error incurred by using only the first few terms. Methods are suggested for the experimental determination of analytic representations for given systems.

1. INTRODUCTION

1.1 NONLINEAR SYSTEMS

At the present time the most useful methods for mathematical analysis and design of electrical systems are based on the theory of linear systems. The techniques of analysis and design of linear systems have been well developed, and they are used not only for perfectly linear systems but also for almost linear systems.

Many communication and control devices are not nearly linear. Sometimes nonlinearity is essential to the operation of a device, sometimes it is undesirable but unavoidable, and sometimes a nonlinear component, although it is not essential, may give better results than any linear component that might be used in its place. Sometimes nonlinearity is avoided, not because it would have an undesired effect in practice, but simply because its effect cannot be computed. There has therefore been an increasing effort to develop methods of analysis and design for nonlinear devices.

It is appropriate to note here the relation between linear and nonlinear systems. A nonlinear system can be almost linear, but there is no such thing as a linear system that is almost nonlinear. The linear case is a limiting case of nonlinearity, and it is an especially simple, not an especially difficult, limiting case.

We should expect, therefore, that any theory or technique that is adequate for general nonlinear systems must be equally adequate for linear systems. The word "nonlinear" is appropriate only to special techniques; a general theory, applicable to both linear and nonlinear systems, should not be called "nonlinear," but "general." However, the designation "nonlinear" will be used in this report to indicate the breadth of the theory, with the understanding that it is not to be interpreted literally as excluding the special linear case.

1.2 HISTORICAL BACKGROUND

Most of the effort to develop techniques of nonlinear system analysis has been primarily associated with a number of Russian schools. In this connection Poincaré, although he was not a Russian, must be mentioned, as well as Liapounoff, Andronov and Chaikin, Kryloff and Bogoliuboff. A great deal of this work was summarized by Minorsky (1) and published in 1947. This earlier research was directed principally toward the solution of nonlinear differential equations and the investigation of the properties of their solutions. Fruitful as this work was, its scope is limited, and it has played no part in the author's research.

The author's research is based on the representation of nonlinear systems by expressing the output directly in terms of the input. The roots of this approach might be historically traced to Volterra (2), who included a theory of analytic functionals in his "Leçons sur les fonctions de lignes" in 1913. In 1923, Wiener (3) brought the theory of Brownian motion to bear on the problem of defining an integral over a space of functions, and included a discussion of the average of an analytic functional. In 1942, Wiener brought Brownian motion and analytic functionals together again (4). The later paper

contains the first use, in the representation of nonlinear systems, of the formula that forms the basis of Section IV of this report, and, in fact, it seems to be the first attempt at a general formula for the representation of nonlinear systems. Some other work along the same lines was done more recently by Ikehara (5) in 1951, and by Deutsch (6) in 1955.

In recent years Wiener developed a general representation method for nonlinear systems that is based on the properties of Brownian motion, but does not employ the formula that he used in 1942. This theory differs from the 1942 report in that it attacks the general nonlinear problem rather than the specific problem of noise in a particular class of systems. The method has been presented in unpublished lectures and described, although not in its most recent form, by Booton (7) and by Bose (8, 9). Theoretical approaches related to this method have been developed by Singleton (10) and by Bose (9). The representation formula developed by Zadeh (11) is similar in its basic orientation.

1.3 SYSTEMS AND FUNCTIONS

One of the central problems in the analysis of nonlinear systems is the finding of a good representation formula. Such a formula must be able to represent, either exactly or with arbitrarily small error, a large class of systems; and it must also be convenient for use in calculations involving systems.

There is, however, a representation problem in a more fundamental sense. It is necessary to reduce the idea of a nonlinear system to more fundamental concepts. This leads to an abstract representation, whose generality is not limited by any concession to computational convenience. With such a representation at hand, representation formulas designed for computational needs can be more easily apprehended.

This abstract representation is found in the general concept of a function. A function, abstractly defined, is a relation between two sets of objects, called the domain and the range of the function, which assigns to every object in the domain a corresponding object in the range, with every object in the range assigned to at least one object in the domain. It may be said that a function is any relation of the form "plug in x, out comes y"; the set of all x that can be plugged in is the domain, and the set of all y that can come out is the range.

This definition implies no restriction on the nature of the objects x and y. We may have, for example, an amplifier chassis with an empty tube socket; every tube that can be inserted in the socket will give us a different amplifier. Therefore, we have a function; the domain of this function is the set of all tubes that can be inserted in the socket, and the range is the set of all amplifiers that can thus be obtained.

We are most familiar with functions whose domain and range are sets of real numbers. Such functions are called "real-valued functions of a real variable"; for convenience, we shall call them "real functions." In general, any function whose range is a set of real numbers is called a "real-valued function." A real function is usually represented by a letter, such as f. The equation $y = f(x)$ means that y is the element of the range which f assigns to the element x of the domain. Note that f(x) is not a function,

but a value of the function f; that is, an element of the range.

A nonlinear system with one input and one output is a function according to this definition. For every input in the set of inputs that the system is designed to accept, the system produces a corresponding output. These inputs and outputs can themselves be represented by functions. If we assume that the inputs and outputs are electric signals, they can be described by real functions: To every real number t there is assigned a corresponding real number f(t) that represents the value of the input or output at time t. A nonlinear system can therefore be represented by a function whose domain and range are sets of real functions. Although such a function is conventionally called an "operator" or a "transformation," it will be referred to in this report as a "hyperfunction" to emphasize the fact that it is a function. A hyperfunction (or the system it represents) will be denoted by a capital script* letter; the equation $g = \underline{H}(f)$ states that g is the real function that represents the output of the system $\underline{H}$ when the input is the signal represented by the real function f.

Most of the discussion in the following sections deals specifically with time-invariant systems. Such systems can be represented by a kind of function that is simpler than a hyperfunction – a function whose domain is a set of real functions and whose range is a set of real numbers. Such functions are conventionally called "functionals."

The argument will be simpler if we consider only physically realizable systems, that is, systems in which the value of the output at any time does not depend on future values of the input. If the system $\underline{H}$ is physically realizable and time-invariant, then the output at a particular time t can be determined without knowing either the value of t or the time at which each value of the input occurred; it is sufficient to specify, for every nonnegative number τ, what the value of the input was τ seconds ago. This input data can be expressed by the real function u, $u(\tau) = f(t-\tau)$ for $\tau \geq 0$, where f represents the input in the usual form. To each function u there corresponds a unique real number $\underline{h}(u)$, with the property that the value of the output of the system is $\underline{h}(u)$ whenever the past of the input is represented by u. The function $\underline{h}$ is a functional according to the definition given. For a specified input f, the function u will be different for different t and will be designated as u_t if t is to be specified; as t changes, u_t changes, and the value of the output changes with it. If the system $\underline{H}$ is not physically realizable, but is still time-invariant, the only change that is necessary in this argument is to define $u(\tau)$ for all τ, negative as well as positive.

For the most part, we shall consider systems for bounded inputs only. A real function f, representing an input, will be called bounded(R) if $|f(t)| \leq R$ for all t. The set of all real functions u, $u(\tau)$ defined for $\tau \geq 0$, that are bounded(R), will be called PBI(R). (PBI stands for Past of Bounded Input.) All these real functions will be assumed to be

*Editor's note: With the permission of the author, the script letters originally used (i.e., $\mathscr{H}$, $\mathscr{h}$, $\mathscr{I}$, etc.) have been replaced with the corresponding typed letter and identified by an underline.

Lebesgue measurable; in practice this is no restriction, since some tricky mathematical work is required to prove the existence of functions that are not Lebesgue measurable. Such "improper functions" as impulses or infinite-bandwidth white noise are not really functions, and thus their measurability is questionable, but they are excluded from consideration as possible inputs on the ground that they are not bounded.

We shall always consider two real functions f and g to be equivalent if

$$\int_a^b [f(x) - g(x)]\, dx = 0 \qquad (1)$$

for all real numbers a and b, since two such functions are indistinguishable by any physical measurement process.

1.4 REPRESENTATION OF FUNCTIONS

The central problem of computationally convenient representation can now be treated with some perspective. We have to find convenient representations for certain kinds of functions, namely, functionals and hyperfunctions.

Suitable methods can be derived by generalizing the familiar methods used for the representation of real functions. These include: (a) miscellaneous designations for special functions, e.g., algebraic, trigonometric; (b) implicit functions; (c) tables of values; (d) polynomials, including power series; and (e) expansions in series of orthogonal functions. The last three are methods of approximate representation, or representation as a limit of successive approximations. All the methods mentioned in section 1.2 are particular forms of generalizations of these methods.

Several classes of specially designated systems, that is, method (a) of the preceding paragraph, are already well known. Perhaps the most important is the class of linear systems, whose special representation by means of the convolution integral has been found particularly convenient. No-memory systems (the value of whose output at any time depends only on the value of the input at that time), differential operators (not differential equations, but such direct statements as "the output is the derivative of the input"), and integral operators [among which are the integral operators of Zadeh (11)] are also specially represented.

An implicit function, method (b), is an equation that does not give f(x) directly in terms of x, but specifies a condition jointly on x and f(x) so that for any x there is a value for f(x) that will satisfy the condition. A differential equation is exactly this sort of condition: given any input f, it is necessary to go through a process called "solving the differential equation" in order to obtain the output g. The methods devised by the Russian schools for obtaining such solutions are all special methods, restricted to certain kinds of equations and certain kinds of inputs, just as the methods of solution for implicit real functions are all special methods.

Approximation methods (c), (d), and (e) are more generally applicable, since they

do not require special forms for the representations of the functions, although they do require that some conditions be satisfied. For the methods that will be discussed, a sufficient condition for arbitrarily close approximation is that the function that is to be represented be continuous and have a compact domain. The interpretation of these conditions for systems will be discussed in Section II. The methods themselves will now be briefly described.

A table of values (c) is conceived of here as being used in the simplest possible manner, that is, without interpolation. In the construction of the table a finite set of x_i is selected from the domain, and for each selected x_i the corresponding $f(x_i)$ is tabulated. In the use of the table, for any given x the nearest tabulated x_i is selected and the corresponding $f(x_i)$ is taken as an approximation to f(x). Owing to the way in which the table is used, its construction can be modified. First, since each tabulated x_i is actually used to represent a set of neighboring x's, the entry in the table may be a designation for this set instead of a particular x_i in the set. Second, since each tabulated $f(x_i)$ is used to approximate a set of f(x), the tabulated value need not be a particular $f(x_i)$ but may be simply a value that is representative of this set of f(x). Either of these schemes can be translated into a method for the approximate representation of functionals by replacing x by a and f by h. The modified scheme is then a general description of Singleton's method for approximating nonlinear systems by finite-state transducers (10). Bose's method of representation (9) also employs the device of a finite table of values. Another method involving tables of values is given in Section III.

An abstract definition of a polynomial (d) will be given in Section IV, as well as the particular form of polynomial representation that was also used by Wiener (4), Ikehara (5), and Deutsch (6). For our present purpose, it is sufficient to note that the sum of a constant, a linear system, and products of linear systems (obtained by using the same input for all systems and multiplying and adding the outputs) is a polynomial system. The formula used in Section IV is somewhat more general than this, and has been found to be convenient for the computations that are required in systems analysis.

Expansions in orthogonal functions (e) will be discussed in Section V. These methods give promise of being convenient for the measurement of nonlinear systems in the laboratory, and their advantages can be combined with the computational convenience of polynomials by using expansions in orthogonal polynomials. The generalization of these methods from real functions to systems is quite interesting. As we know from the theory of real functions, expansion of a function in orthogonal functions involves integration over the domain of the function. Integration over a set of real numbers is a familiar process, but how can we integrate over a set of functions? Definition of such integrals was the essential problem that Wiener (3) attacked in 1923, and at that time it was a difficult problem. Now, however, probability theory offers a solution: on a statistical ensemble of functions, which is just a set of functions with probabilities defined on it, an ensemble average (expectation) is equivalent to an integral. This is the essential reason for the introduction of probability in the methods of Booton (7) and Bose (8), as well as in the

method of Wiener described by Booton and Bose; these can be [illegible] ethods of expanding a nonlinear system in a series o. orthogonal syste[illegible].

The following sections discuss some example[illegible] of approxim[illegible] s, t[illegible]r applications, and some sufficient conditions for their applicability [illegible] eals with conditions of approximability, and the next three sections are dev[illegible] e general methods of approximation.

II. APPROXIMATIONS TO NONLINEAR SYSTEMS

2.1 TOPOLOGY AND APPROXIMATIONS

The primary aim of this section is to establish some sufficient conditions for the approximability of a nonlinear system by the methods that will be described in subsequent sections. The most important results of this section are summarized in section 2.7.

The theorems that will be developed are essentially theorems of analysis; in fact, one theorem of analysis, the Stone-Weierstrass theorem, will be quoted and used without proof. Most of the mathematical ideas can be found, in the restricted context of real functions, in Rudin's "Principles of Mathematical Analysis" (12); the Stone-Weierstrass theorem that he proved is applicable to our purpose. For a discussion of analysis in a more general setting, especially for a general definition of a topological space, and for a more appropriate definition of a compact set than is given in Rudin, reference can be made to Hille's "Functional Analysis and Semi-Groups" (13).

One way in which a topology may be rigorously defined is in terms of neighborhoods. A topological space is a set of objects x in which certain subsets N(x) are designated as neighborhoods of specific objects x. [Usually, there is an infinity of objects x and, for each x, an infinity of N(x).] These neighborhoods satisfy certain conditions that constitute the postulates of topology: first, every x has at least one N(x), and every N(x) contains x; second, if $N_A(x)$ and $N_B(x)$ are two neighborhoods of the same object x, there is an $N_C(x)$ with the property that any object in $N_C(x)$ is also in both $N_A(x)$ and $N_B(x)$; third, for any object y contained in any neighborhood $N_A(x)$ there is an $N_B(y)$ with the property that any object in $N_B(y)$ is also in $N_A(x)$. (Conventionally, the objects in a topological space are called "points." This term will not be used in this report because it suggests a very restricted interpretation of topology.)

It will now be shown that topology as just defined is a mathematical analogue of the engineering idea of approximation. Practically, approximations occur when we consider some object (e.g., a number, a position in space, a resistor, a signal, a system) that is to be used for some purpose, and want to know what other objects are sufficiently similar to it to be used for the same purpose. We thus define a criterion of approximation to this object, and consider the set of all objects that, by this criterion, are good approximations to it. It will be shown that these approximation sets, as neighborhoods, satisfy the postulates of topology that have been given.

First, every object considered by engineers is usable for some purpose, and thus at least one neighborhood is defined for it; and for any purpose an object is always a good approximation to itself. Second, if an object x can be used for two purposes A and B, two neighborhoods $N_A(x)$ and $N_B(x)$ thus being defined, we consider purpose C as the requirement of being sufficiently similar to x to satisfy both purposes A and B; thus defines a neighborhood $N_C(x)$ with the property that every object in $N_C(x)$ is also in both $N_A(x)$ and $N_B(x)$. Third, given x and some $N_A(x)$, and any y in $N_A(x)$, we can consider purpose F for y as that of substituting for x in the fulfillment of purpose A, and can

define $N_B(y)$ as the set of all objects that are sufficiently similar to y to serve this purpose; then every object in $N_B(y)$ is also in $N_A(x)$.

These arguments may seem trivial and pointless; actually they establish the relation between topology and approximations and make the topological foundations of analysis, and all the theorems that follow from them, applicable to engineering.

For any set of objects, different classes of approximation criteria can often be used, with the result that different sets of neighborhoods and different topologies are obtained. However, different sets of neighborhoods do not always lead to different topologies. Two sets of neighborhoods are said to be bases of the same topology if every neighborhood in each set contains at least one neighborhood from the other set. This is because the closed sets, open sets, compact sets, and continuous functions (defined in section 2.2) are the same for both sets of neighborhoods.

On a space of real numbers, a neighborhood of a number x is defined by the property that y is in $N_\epsilon(x)$ if the magnitude of the difference between x and y is less than ϵ. In the uniform topology on a space of real functions, g is in $N_\epsilon(f)$ if, for every real number t, the magnitude of the difference between f(t) and g(t) is less than ϵ; a similar condition defines the uniform topology on a space of functionals. On a space of hyperfunctions (or, equivalently, systems), we define the uniform topology by the statement that $\underline{K}$ is in $N_\epsilon(\underline{H})$ if, for every input f, at every time t, the magnitude of the difference of the values of $\underline{K}(f)$ and $\underline{H}(f)$ is less than ϵ; or, equivalently, $\underline{K}(f)$ is in $N_\epsilon(\underline{H}(f))$ for every f. A different topology on a space of real functions will be defined in section 2.4.

2.2 SOME TOPOLOGICAL CONCEPTS

A number of topological ideas that are to be used in the discussion of approximations to nonlinear systems will now be defined. We begin by defining open and closed sets, in spite of the fact that we shall make no use of them, not only because mathematical tradition seems to demand it, but also because many writers define topology in terms of open sets, rather than in terms of neighborhoods.

An open set is a set with the property that every object in the set has at least one neighborhood that is contained in the set. A closed set is a set whose complement — the set of all objects in the space that are not in the set — is open. An equivalent definition is that a closed set is a set that contains all its limit points. When a topological space is defined in terms of open sets, neighborhoods are usually defined by calling every open set a neighborhood of every object that it contains.

A limit point of a set A is an object (which may or may not be in A) every neighborhood of which contains at least one object in A other than x. In other words, a limit point of A is an object that can be approximated arbitrarily closely (i.e., under any criterion of approximation) by objects, other than itself, in A.

The closure of a set is the set of all objects that are either in the set or are limit points of the set (or both). In other words, the closure of a set A is the set of all objects

that can be approximated arbitrarily closely by objects in A. In the application of this concept we shall consider the closure of the set of all systems that can be exactly represented by some method; the closure will be the set of all systems that can be represented by this method, either exactly or with arbitrarily small error.

A compact set is defined as follows. A collection of neighborhoods is said to cover a set A if every object in A is in at least one of the neighborhoods in this collection. A set is called compact if every collection of neighborhoods that covers it includes a finite subcollection that also covers the set. If we define a criterion of approximation for every object in the set A, by choosing a neighborhood for every object in A, this collection of neighborhoods covers A; and if A is compact we can select a finite set of objects in A with the property that every object in A is in the chosen neighborhood of at least one of the selected objects. The importance of this property can be indicated by interpreting neighborhoods in a slightly different way; that is, by considering a neighborhood of an object as a set of objects that x can approximate, instead of a set of objects that can approximate x. (These interpretations are equivalent if the approximation criterion has the property that x approximates y whenever y approximates x.) Then a compact set is one such that, for any predetermined criterion of approximation, can be approximated by a finite subset of itself.

Topology is combined with the abstract idea of a function in the definition of a continuous function. Suppose the range and domain of a function f are both topological spaces; f is said to be continuous if for every x in the domain, and for any neighborhood $N_A(f(x))$ of the corresponding f(x), there is a neighborhood $N_B(x)$ with the property that whenever y is in $N_B(x)$, f(y) is in $N_A(f(x))$. (Note that N_A and N_B are neighborhoods in different spaces.) This is a precise statement of the imprecise idea that a continuous function is one whose value does not change abruptly; it implies that any approximation criterion in the range can be satisfied by an appropriate criterion of approximation in the domain.

2.3 TWO THEOREMS OF APPROXIMATION

In terms of the concepts previously defined, two important theorems on approximation of functions can be stated. These theorems will be applied to nonlinear systems in section 2.4.

The first is a theorem on representation by tables of values. Let f be a continuous function with a compact domain. Let a neighborhood $N_A(y)$ be chosen for every y in the range. Then there is a finite set of objects x_i in the domain, and for each x_i a neighborhood $N_B(x_i)$, such that every x in the domain is in at least one $N_B(x_i)$, and, whenever x is in $N_B(x_i)$, f(x) is in $N_A(f(x_i))$.

To apply this theorem, we consider functions whose ranges are sets of real numbers or real functions, such as functionals or hyperfunctions. We choose a positive real

number ϵ as the tolerance for a criterion of approximation, and define neighborhoods in the range, as in section 2.1. Suppose some topology is also defined in the domain, and that with these two topologies the function f is continuous. Then we can select a finite set of objects x_i and neighborhoods $N(x_i)$, as indicated in the theorem; we construct a table of these x_i and the corresponding $f(x_i)$. Then, for any x in the domain, we can find in the table an x_i with the property that x is in $N(x_i)$, and the tabulated $f(x_i)$ will differ from f(x) by less than ϵ.

The proof of this theorem is quite simple. Since the function is continuous, there is a neighborhood $N_B(x)$ for every x with the property that, if x' is in $N_B(x)$, f(x') is in $N_A(f(x))$. The collection of all these neighborhoods $N_B(x)$ covers the domain, and, since the domain is compact, there is a finite set of x_i with the property that the collection of $N_B(x_i)$ also covers the domain. This set of x_i fulfills the conditions stated in the theorem, and the theorem is thus proved. Incidentally, we have also proved that if a function is continuous and its domain is compact then its range is also compact.

The second theorem to be stated here is the Stone-Weierstrass theorem; in effect, it is a theorem on the approximation of functions by polynomials. It is restricted to real-valued functions although the nature of the domain is not restricted. It is similar to the first theorem in that we assume that the function to be approximated is continuous with compact domain. The statement of this theorem must be preceded by some preliminary definitions.

If f, f_1, and f_2 are functions with the same domain and A is a real number, then $f = f_1 + f_2$ if $f(x) = f_1(x) + f_2(x)$ for every x in the domain, $f = f_1 f_2$ if $f(x) = f_1(x) f_2(x)$ for every x in the domain, and $f = Af_1$ if $f(x) = Af_1(x)$ for every x in the domain. These definitions, although obvious, are logically nontrivial.

An algebra of functions is a set of functions, all of which have the same domain, with the property that for every f and g in the set and for every real number A, the functions f+g, fg, and Af are also in the set.

An algebra of functions is said to separate points if, for every pair of objects x and y in their domain and every pair of real numbers A and B, there is a function f in the algebra with the property that $f(x) = A$ and $f(y) = B$.

The Stone-Weierstrass theorem states that if an algebra of real-valued continuous functions has a compact domain and separates points, then the closure of the algebra, in the uniform topology, is the set of all continuous real-valued functions with that domain; i.e., for any continuous real-valued function f with that domain, and any positive number ϵ, there is a function g in the algebra such that $|f(x) - g(x)| < \epsilon$ for every x in the domain.

The proof of this theorem has been given by Rudin (12); it is too involved to be repeated here. Although the context of Rudin's proof may suggest that the theorem concerns only functions of a real variable, the same proof is valid for compact domains in the most general topological spaces.

2.4 A SPECIAL TOPOLOGICAL SPACE

The approximation theorems of section 2.3 will now be applied to nonlinear systems. Specifically, since it was shown in Section I that a time-invariant system can be represented by a functional, they will be applied to functionals.

The theorems indicate that a sufficient condition for a function to be approximable is that it be continuous and have a compact domain. These properties depend upon the topologies on the domain and the range. On the range (which is a set of real numbers) there is only one useful topology; but on the domain (which is a set of real functions) a fairly wide choice of topologies is possible. The practical meaning of the theorems depends upon the topology that is used on the domain; but if the theorems are to have any practical meaning at all, the topology that is used must imply both a physically meaningful definition of continuity and the existence of physically significant compact sets.

A topology that meets these requirements has been found. (Further research might reveal others.) On the space PBI(R), which was defined in Section I as the set of all real functions u for which $|u(\tau)| \leq R$ for all τ in the domain $0 \leq \tau < \infty$, neighborhoods $N_{T,\delta}(u)$ are defined as follows: v is in $N_{T,\delta}(u)$, $T > 0$, $\delta > 0$, if and only if

$$\left| \int_0^x [u(\tau) - v(\tau)]\, d\tau \right| < \delta \qquad (2)$$

for all x in the interval $0 \leq x \leq T$. The topology defined by these neighborhoods will be called the RTI (Recent Time Integral) topology.

This condition may be alternatively expressed by defining the functions U and V,

$$U(x) = \int_0^x u(\tau)\, d\tau, \qquad V(x) = \int_0^x v(\tau)\, d\tau \qquad (3)$$

Then v is in $N_{T,\delta}(u)$ if and only if the magnitude of the difference between U(x) and V(x) is less than δ for every x in the interval $0 \leq x \leq T$. Note that if v is in $N_{T,\delta}(u)$, then u is in $N_{T,\delta}(v)$, and vice versa. It will be seen that for v to be in $N_{T,\delta}(u)$ no condition need be imposed on the values of these functions for $\tau > T$, although for $\tau \leq T$ the difference between $u(\tau)$ and $v(\tau)$ need not remain small, but may alternate rapidly between large positive and negative values.

It will be shown in the next section that the space PBI(R), for any R, is compact in the RTI topology. We therefore consider a functional $\underline{h}$ whose domain is the space PBI(R). This functional is continuous if, for any positive number ϵ, there exist positive numbers T (sufficiently large) and δ (sufficiently small) such that if v is in $N_{T,\delta}(u)$ (and v and u are both in PBI(R)), then $|\underline{h}(u) - \underline{h}(v)| < \epsilon$. The functional will then be called continuous(R), and the time-invariant system $\underline{H}$ that it represents will also be called continuous(R). Any system referred to as continuous is understood to be time-invariant.

In the representation of systems by functionals, the function u represents the past of the input. We may therefore interpret continuity for nonlinear systems (with respect to the RTI topology) by the statement that a system is continuous(R) if, for all inputs that are bounded(R), the value of the output is relatively insensitive to small changes in the input, to rapid fluctuations (high frequencies) in the input, and to the remote past of the input.

It follows from the first theorem of section 2.3 that a system $\underline{H}$ that is continuous(R) can be represented with any desired accuracy by a finite table of values, since the functional that represents it is a continuous function with a compact domain. Let any tolerance ϵ be given; then T and δ are determined according to the continuity condition, a finite set of real functions u_i is selected with the property that the collection of neighborhoods $N_{T,\delta}(u_i)$ covers PBI(R), and these real functions u_i are tabulated with the corresponding values $\underline{h}(u_i)$.

It will be shown in section 2.5 that a time-invariant linear system is continuous(R) for any R if and only if its impulse response is Lebesgue integrable; this is roughly equivalent to the condition that its transients be damped and that its impulse response involve no impulses. The set of all such linear systems, all products of these systems, all constant-output systems, and all sums of these, is an algebra. The functionals that represent them constitute an algebra of continuous functionals. It is easy to show that this algebra separates points. The Stone-Weierstrass theorem then implies that any functional that is continuous(R), with domain PBI(R), can be approximately represented, with arbitrarily small error, by a functional chosen from this algebra. Hence, any system that is continuous(R) can be approximated arbitrarily closely in polynomial form.

2.5 CONTINUITY AND COMPACTNESS IN THE RECENT TIME INTERVAL (RTI) TOPOLOGY

This section is devoted to proofs of two statements made in section 2.4: that a time-invariant linear system is continuous(R), for any R, if and only if its impulse response is Lebesgue integrable; and that the space PBI(R) is compact in the RTI topology.

The theorem on continuity of a linear system will be proved first. A time-invariant linear system is represented by the functional $\underline{h}$, defined by

$$\underline{h}(u) = \int_0^\infty h(\tau)\, u(\tau)\, d\tau \tag{4}$$

where h is the impulse response of the system. Suppose that h is not Lebesgue integrable; this may be so either because

$$\int_0^\infty |h(\tau)|\, d\tau = \infty \tag{5}$$

(i.e., h is not absolutely integrable), or because the integral of h is so defined that it is not equal to the Lebesgue integral (e.g., h involves impulses). It will be shown in each of these two cases that $\underline{h}$ is not continuous(R) for any R.

Suppose h is not absolutely integrable. Choose $\epsilon > 0$, and try to find a T and δ with the property that if v is in $N_{T,\delta}(u)$ then $|\underline{h}(u) - \underline{h}(v)| < \epsilon$. But we can choose u and v so that $v(\tau) - u(\tau)$ has a constant magnitude less than δ/T, and an algebraic sign that is always equal to the sign of $h(\tau)$; then v is in $N_{T,\delta}(u)$, but the difference between $\underline{h}(u)$ and $\underline{h}(v)$ is infinite.

Now suppose that h contains an impulse of value A (i.e., A is the integral of the impulse) at τ_o. Choose ϵ less than $|2AR|$ and try to find a corresponding T and δ. But if we choose v and u so that their values are equal except on a small interval that contains τ_o, we can have v in $N_{T,\delta}(u)$ by making this interval small enough and still have $|u(\tau_o) - v(\tau_o)| = 2R$, with the result that $|\underline{h}(u) - \underline{h}(v)| = |2AR| > \epsilon$. A similar argument holds whenever the impulse response is absolutely integrable, but not Lebesgue integrable, since in that case the indefinite integral of the impulse response is not absolutely continuous.

Now suppose that the impulse response h is Lebesgue integrable. We prove that $\underline{h}$ is a continuous functional. We consider the domain of the functional to be PBI(R) for any chosen R, and choose any $\epsilon > 0$. Let $p = \epsilon/4R$. Now construct a step-function h_p – a real function whose value is constant on each of n bounded intervals and is zero outside them – so that

$$\int_0^\infty |h_p(\tau) - h(\tau)|\, d\tau \leq p \tag{6}$$

The existence of such a step-function can be proved from the fundamental definitions of the Lebesgue integral; it is obvious if h is continuous. There is a number M with the property that $|h_p(\tau)| \leq M$ for all τ, and a number T with the property that $h_p(\tau) = 0$ for all $\tau > T$. Let $\delta = \epsilon/6nM$.

Let v be in $N_{T,\delta}(u)$ for these values of T and δ. Then for any one of the n intervals, say $a \leq \tau \leq b$, we have

$$\left|\int_a^b h_p(\tau)\, u(\tau)\, d\tau - \int_a^b h_p(\tau)\, v(\tau)\, d\tau\right| = \left|\int_a^b h_p(\tau)[u(\tau) - v(\tau)]\, d\tau\right| \tag{7}$$

$$\leq M\left|\int_a^b [u(\tau) - v(\tau)]\, d\tau\right| \leq 2M\delta = \epsilon/3n$$

Since there are n intervals,

$$|\underline{h}_p(u) - \underline{h}_p(v)| \leq \epsilon/3 \tag{8}$$

Also, since u is bounded and h_p approximates h,

$$|\underline{h}(u) - \underline{h}_p(u)| \leq pR = \epsilon/4 \tag{9}$$

so that as we pass from $\underline{h}(u)$ to $\underline{h}_p(u)$ to $\underline{h}_p(v)$ to $\underline{h}(v)$ the changes are all small; and

$$|\underline{h}(u) - \underline{h}(v)| \leq \epsilon/4 + \epsilon/3 + \epsilon/4 < \epsilon \tag{10}$$

and the continuity of $\underline{h}$ is proved.

We turn now to proving the statement that the space PBI(R) is compact in the RTI topology. The proof is accomplished by reductio ad absurdum; we assume that the space is not compact and derive two contradictory conclusions.

We begin by constructing a special set of functions $v_{n,k}$ in the space PBI(R). One set of functions $v_{n,k}$ for a given n is constructed by dividing the interval $0 \leq \tau \leq 2^n$ into 2^{2n} subintervals of length 2^{-n}; the value of each function $v_{n,k}$ is constant on each subinterval and equal to either R or -R, and is zero for $\tau > 2^n$. For each n, the index k will therefore run from 1 to $2^{2^{2n}}$. Since the number of these functions for each n is finite, all these functions, for all n, can be arranged in an infinite sequence in order of increasing n.

With each $v_{n,k}$ we associate the particular neighborhood $N^*(v_{n,k}) = N_{T,\delta}(v_{n,k})$ defined by $T = 2^n$ and $\delta = 2^{1-n}R$. It will now be shown that for any n the collection of neighborhoods $N^*(v_{n,k})$ covers PBI(R). This is equivalent to showing that for any u in PBI(R) and for any integer n we can construct a function $v_{n,k}$ in such a way that the magnitude of

$$E(x) = \int_0^x [v_{n,k}(\tau) - u(\tau)]\, d\tau \tag{11}$$

is less than $2^{1-n}R$ for every x in the interval $0 \leq x \leq 2^n$. We construct such a function $v_{n,k}$, starting from $\tau = 0$, by the following procedure. Suppose that $v_{n,k}(\tau)$ has been determined on the interval $0 \leq \tau \leq \tau_o$. We shall decide whether the value of $v_{n,k}$ is to be R or -R on the interval $\tau_o < \tau \leq \tau_o + 2^{-n}$. According to our decision, $E(\tau)$ will either increase or decrease monotonically on this interval to a value

$$E\left(\tau_o + 2^{-n}\right) = E(\tau_o) - \int_{\tau_o}^{\tau_o + 2^{-n}} u(\tau)\, d\tau \pm 2^{-n}R \tag{12}$$

Since

$$\left| E(\tau_o) - \int_{\tau_o}^{\tau_o + 2^{-n}} u(\tau)\, d\tau \right| < 2^{1-n}R + 2^{-n}R = 3 \cdot 2^{-n}R \tag{13}$$

at least one of these alternatives will make the magnitude of $E\left(\tau_o + 2^{-n}\right)$ less than $2^{1-n}R$. Since the magnitude of $E(\tau_o)$ is also less than $2^{1-n}R$, and E is monotonic on this

interval, the same bound holds everywhere on the interval. We can continue this construction over the entire length of the interval $0 \leq \tau \leq 2^n$, and prove that for any n the collection of $N^*(v_{n,k})$ covers PBI(R).

Now suppose that for each u in PBI(R) we specify a neighborhood $N'(u)$; then the collection of all $N'(u)$ covers PBI(R), and, since we assume that PBI(R) is not compact, we assume that these $N'(u)$ have been so chosen that no finite subcollection of these neighborhoods covers PBI(R). We can prove at the outset that there is a countable subcollection that covers PBI(R) – that is, a collection that can be ordered in a sequence. For each u, $N'(u) = N_{T,\delta}(u)$ for some T and δ. Choose n so large that $2^n > T$ and $2^{1-n} < \delta/2$. There is a $v_{n,k}$ with the property that $N^*(v_{n,k})$ contains u, and this $N^*(v_{n,k})$ is contained in $N'(u)$. Thus, assigning some $N^*(v_{n,k})$ to every u, we obtain a subcollection of the infinite sequence of $N^*(v_{n,k})$. This subcollection covers PBI(R) and can be arranged in a sequence, and each neighborhood is contained in at least one $N'(u)$. For each, pick one of the $N'(u)$ that contains it. We thus obtain a sequence $N'(u_m)$, $m = 1,2,\ldots$, which covers PBI(R).

Now choose $m_1 = 1$. No finite collection of $N'(u)$ covers PBI(R), so there is some function w_1 in PBI(R) that is not in $N'(u_1)$; but, since the collection of $N'(u_m)$ covers PBI(R), there is an m_2 with the property that $N'(u_{m_2})$ contains w_1. Consider next the collection of $N'(u_m)$, $m = 1,2,\ldots,m_2$. This collection does not cover PBI(R), so there is a w_2 that is not in any of these neighborhoods, but there is an m_3 with the property that w_2 is in $N'(u_{m_3})$. Proceed, in this manner, to construct an infinite sequence of w_i, with the property that no finite collection of the neighborhoods $N'(u_m)$ contains more than a finite number of them.

We shall now contradict this conclusion by proving that for any sequence of functions w_i in PBI(R) there is at least one neighborhood $N'(u_m)$ that contains an infinite number of w_i. Consider the collection of $N^*(v_{n,k})$ for $n = 2$. This collection covers PBI(R) and contains a finite number of neighborhoods, so that at least one of them, which we shall call $N^*(v_2)$, contains an infinity of w_i. If we consider only the w_i that are contained in $N^*(v_2)$, the collection of $N^*(v_{n,k})$ for $n = 3$ must contain at least one $N^*(v_3)$ that contains an infinity of these w_i. We proceed, in this way, to construct a sequence of neighborhoods $N^*(v_j)$, each of which contains an infinity of w_i, including any w_i that is contained in any $N^*(v_j)$ that follows it in the sequence.

We now define the functions V_j,

$$V_j(x) = \int_0^x v_j(\tau)\, d\tau \tag{14}$$

It can be shown that for every x the sequence of numbers $V_j(x)$ forms a convergent sequence, so these functions V_j have a limit function V_o. It can also be shown that V_o is absolutely continuous, and that the magnitude of its derivative does not exceed R, so there is a function v_o in PBI(R) with the property that

$$V_o(x) = \int_0^x v_o(\tau)\, d\tau \qquad (15)$$

It will be seen that every neighborhood of v_o contains an infinity of w_i. Since there is at least one $N'(u_m)$ that contains v_o, and every neighborhood that contains v_o contains a neighborhood of v_o, the $N'(u_m)$ contains an infinity of w_i. This statement contradicts our previous conclusion. The assumption that PBI(R) is not compact has led to a contradiction, and we conclude that PBI(R) is compact.

2.6 HYSTERESIS

Hysteresis is often mentioned as a typical phenomenon of nonlinear systems. Examination of methods of approximate representation of nonlinear systems leads to a general impression that these methods fail when hysteresis occurs. However, a general definition of hysteresis is necessary before this impression can be promoted to the status of a conclusion. Such a definition will be proposed in this section. It will then be shown that systems that exhibit the phenomenon of hysteresis are not continuous, in the sense in which continuity was defined in section 2.4.

The phenomenon of hysteresis is usually understood in terms of a hysteresis loop.

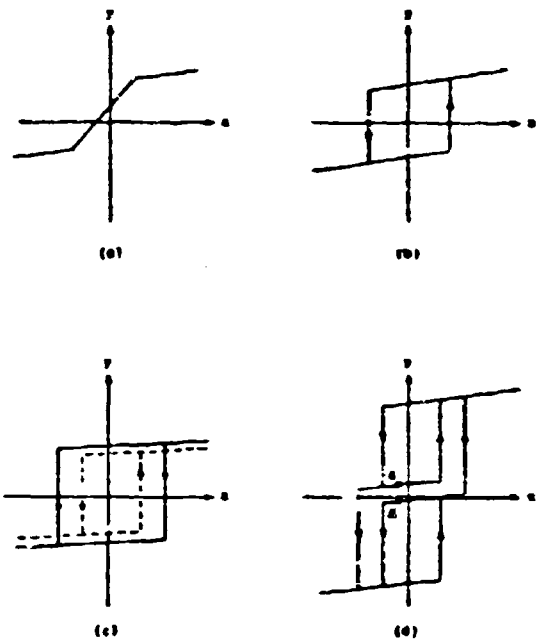

Fig. 1. Hysteresis loops. (a) No hysteresis. (b) With hysteresis. (c) Two systems with hysteresis. (d) Sum of the two systems in (c).

A system that exhibits hysteresis is contrasted with a system in which the instantaneous value of the output, y, is uniquely determined by the instantaneous value of the input, x, as in Fig. 1a. This is a typical no-memory system.

A system with hysteresis is then considered as one that, for a given value of the input, may have one of several different output values. The system is, at any time, in one of several states, the state depending not only on the value of the input, but also on previous states of the system; the state depends, however, only upon the order, and not upon the time, in which these previous states were passed through. A graph representing such a system is shown in Fig. 1b. (For the sake of simplicity, the sides of the loop are made vertical.) When x changes, y, if the graph indicates more than one possible value, assumes the value nearest to the value it had most recently. The vertical parts of the loop can therefore be traversed only in the indicated direction.

In the system of Fig. 1b the state of the system can be determined from the input and output values, but this is not always true. Consider the system of Fig. 1d, which is formed from the two systems of Fig. 1c by applying the same input to each, and adding the outputs. If the two loops of Fig. 1c were of the same height, points A and A' of Fig. 1d would indicate the same input and the same output, but would represent different states.

In such simple cases as these, we can say that the output of a hysteretic system depends on its past history, whereas the output of a nonhysteretic system does not. But this statement is not sufficient in a general context, when the output of a nonhysteretic system (e.g., a linear system) also depends on the past of the input. To define hysteresis for general time-invariant systems, we use two conditions that may be deduced from the graphs of Figs. 1b and 1d, but are meaningful also in a general context.

In a hysteretic system it is possible, first, to specify two inputs, different for $t < 0$ but equal for $t > 0$, for which the difference between the corresponding outputs does not converge to zero as $t \to \infty$. Second, it is possible to specify an input for $t > 0$ in such a way that for any two inputs, arbitrarily different for $t < 0$ but with the specified form for $t > 0$, the difference between the corresponding outputs always converges to zero as $t \to \infty$. (The condition of convergence toward zero is used, instead of equality to zero, to allow for dependence of the output on the finitely remote past of the input.)

Any system that is continuous(R) is an example of a system that does not satisfy the first condition, since, for any ϵ, we specify T as in the definition in section 2.4, and the outputs for $t > T$ must differ by less than ϵ. An example of a system that satisfies the first condition, but not the second, is an ideal linear integrator.

We can prove, by using the principle of superposition, that a linear system must be nonhysteretic. If two inputs are different for $t < 0$ and both are zero for $t > 0$, then adding to each input a specified component for $t > 0$ can only add to each output a component that is the same for both outputs, and cannot change the difference between the outputs. Hence if a linear system satisfies the first condition for hysteresis, it cannot satisfy the second.

A system capable of subharmonic response to a sinusoidal input satisfies the first condition of hysteresis, since two inputs that are equal for $t > 0$ may give rise to subharmonic outputs in different phase. If, in such a system, the output tends to zero when the input becomes and remains zero, the second condition is satisfied and the system is hysteretic.

Since a hysteretic system is not continuous, it cannot be approximated arbitrarily closely, in the uniform topology, by continuous systems. However, for some hysteretic systems and some input ensembles, it may be possible to find continuous systems that approximate the output with arbitrarily small error except for an arbitrarily small probability of large error. The reason is that the second condition in the definition of hysteresis implies that events can occur in the input that make the system forget what happened before them. For some ensembles and inputs there can be a positive probability of such an event occurring in a bounded time interval, and this probability will approach unity as the length of the interval increases; thus a system whose output depends effectively on a finite portion of the past of the input may be made to approximate the output with arbitrarily small probability of large error.

2.7 SUMMARY

The principal ideas and conclusions are outlined in this section. These ideas form the foundation of the methods of representation that will be discussed in subsequent sections.

The input to a system is represented by a real function f; the value of the input at time t is f(t). The output is similarly represented by a real function g. If f is Lebesgue measurable, and $|f(t)| \leq R$ for all t, then f is said to be bounded(R).

A nonlinear system $\underline{H}$ is a function. It assigns to every input f (in a specified set) a corresponding output $g = \underline{H}(f)$.

A time-invariant system $\underline{H}$ can be represented by a functional $\underline{h}$. The value of the output at time t is determined from the function u_t, defined by $u_t(\tau) = f(t-\tau)$, $\tau \geq 0$, then $\underline{h}(u_t) = g(t)$. If f is bounded(R), then u_t is an element of the set of functions called PBI(R).

A topology is a scheme of approximation criteria; a neighborhood is a set of approximations satisfying some criterion. On the space PBI(R) we define the RTI topology by the condition that, for $T > 0$, $\delta > 0$, the function v is in the neighborhood $N_{T,\delta}(u)$ of the function u if and only if

$$\left| \int_0^x [u(\tau) - v(\tau)]\, d\tau \right| < \delta \qquad (16)$$

whenever $0 \leq x \leq T$.

A time-invariant system $\underline{H}$, considered only for inputs that are bounded(R), is said to be continuous(R) if, for any $\epsilon > 0$, there exist $T > 0$, $\delta > 0$ (T sufficiently large, δ sufficiently small) such that if u and v are in PBI(R) and v is in $N_{T,\delta}(u)$, then $|\underline{h}(u) - \underline{h}(v)| < \epsilon$.

If $\underline{H}$ is continuous(R), then for any $\epsilon > 0$ there is a finite set of functions u_i in PBI(R), and a neighborhood $N_{T,\delta}(u_i)$ of each u_i, such that every u in PBI(R) is in at least one of these neighborhoods, and if u is in $N_{T,\delta}(u_i)$, then $|\underline{h}(u) - \underline{h}(u_i)| < \epsilon$.

If $\underline{H}$ is continuous(R), then for any $\epsilon > 0$ there is a polynomial system $\underline{H}_\epsilon$, consisting of a sum of a constant, a linear system with Lebesgue integrable impulse response, and products of such linear systems, such that, for any input that is bounded(R), the values of the outputs of $\underline{H}$ and $\underline{H}_\epsilon$ never differ by more than ϵ.

III. A DIGITAL APPARATUS

THEORY

As an illustration of some of the results of Section II, a description and discussion of a hypothetical apparatus for the analysis and synthesis of nonlinear systems will be presented. The apparatus is designed for the laboratory examination of an unknown system and the synthesis of an approximately equivalent system. It has not been built because it does not appear to be practical, but some of its principles are interesting.

The apparatus is based on the approximate representation of functionals by means of tables of values. To represent a functional $\underline{h}$, we tabulate the values of $\underline{h}(u)$ for a finite number of real functions u. The functions for which values are tabulated in this apparatus are similar to the functions $v_{n,k}$ used in the proof of a theorem in Section II. They are constructed by quantizing time in intervals of equal length q and by making the value of the function constant, and equal to either R or -R, on each interval. We shall call such functions quantized functions. They can be represented as sequences of binary symbols, 1 for value R and 0 for -R.

Given any input function f that is bounded(R), we generate a quantized function f^* to approximate it in the sense that $\left|\int_A^B [f(t) - f^*(t)]\,dt\right|$ is made as small as possible for every A and B. The value of f^* on each q-interval must obviously be determined later than the beginning of that interval, without any knowledge of the values f will have on that interval. We may use a feedback method, and the error of approximation can be determined by means of an integrator. The input to the integrator will be $f - f^*$, and the output will be the error signal e,

$$e(t) = \int_{-\infty}^{t} [f(x) - f^*(x)]\,dx \tag{17}$$

and, at the beginning of each q-interval, we make the value of f^* equal to R on that interval if e(t) is positive, and equal to -R if e(t) is negative; if e(t) = 0 we make an arbitrary choice.

On an interval of length q, f^* will contribute a change of magnitude qR to the value of e, with sign opposite to the sign of e at the beginning of the interval, whereas f will contribute an unpredictable change of magnitude that does not exceed qR, so that the value of e(t) will be kept within the bounds of $\pm 2qR$. Then

$$\left|\int_A^B [f(t) - f^*(t)]\,dt\right| = |e(A) - e(B)| \leq 4qR \tag{18}$$

This error is attributable partly to quantization and partly to the delay in using the values of f to determine the values of f^*. In deriving from f^* an approximation to u, it is therefore desirable to make use of the fact that f^* can be predicted up to the end of the present interval. We define $u_1^*(\tau) = f^*(t^* - \tau)$, where t^* is the end of the q-interval

containing t. This means that $u_t^*(\tau)$ is constant on intervals of length q, starting from $\tau = 0$, and the values of u_t^* on these intervals are the values already determined for f^*, beginning with the most recent. Then for all t and all x,

$$\left| \int_0^x [u_t(\tau) - u_t^*(\tau)]\, d\tau \right| < 4qR \tag{19}$$

Given a system $\underline{H}$ that is continuous(R), and given ϵ as the desired tolerance for the output, determine T and δ to satisfy the continuity definition, and let $q = \delta/4R$. Let n be the smallest integer that is not less than T/q (i.e., the number of intervals of length q needed to cover an interval of length T); u_t^* need be defined only on the first n intervals of length q, and may be taken as zero for $\tau > nq$. Then u_t^* is in $N_{T,\delta}(u_t)$, and therefore $|\underline{h}(u_t^*) - \underline{h}(u_t)| < \epsilon$.

The function u_t^* can be represented as a sequence of n binary digits. It must be one of 2^n possible functions u^*, each of which is represented by a binary number. For each number we tabulate $\underline{h}(u^*)$.

In many cases it is not practical to determine in advance the appropriate T and δ for a given ϵ. In such cases, n and q can be chosen arbitrarily, and if the resulting error turns out to be too large, it can always be made smaller by choosing q smaller and nq larger.

3.2 CONSTRUCTION AND OPERATION

Two devices are conceived for the implementation of this theory, an analysis device (for examining a system in the laboratory) and a synthesis device. The synthesis device would consist of two parts, one a quantizer to determine u_t^*, and the other a storage device that contains previously determined values of $\underline{h}(u^*)$ and produces, at every time t, the value $\underline{h}'$ ... The analysis device generates quantized inputs f, so that, at certain times t, u_t has the form of u^*, and records the values of the output at these times as $\underline{h}(u^*)$.

Figure 2 illustrates the quantizer. The interval length q is determined by the frequency of the pulse generator, and n is the number of stages in the shift register. The shift register holds, at all times, the last n digits of f^*, and delivers these, as u_t^*, to the storage unit, which, in turn, delivers the output. The curves in Fig. 2 were calculated to indicate typical operation of the quantizer.

The analysis device consists of two parts, a generator that generates quantized inputs and a storage unit to record the output values. In using this device, it is not necessary to wait for the system that is being tested to return to its rest state before every sequence of n digits of input because the system is sensitive, within the specified tolerance, only to the last n digits of any long input sequence. We can therefore use a long sequence of digits with the property that every possible sequence of n digits is contained in it. For example, the sequence 01100 contains the two-digit groups 01, 11, 10, 00 in that order; similarly the sequence 0100011101 contains all possible sequences of three

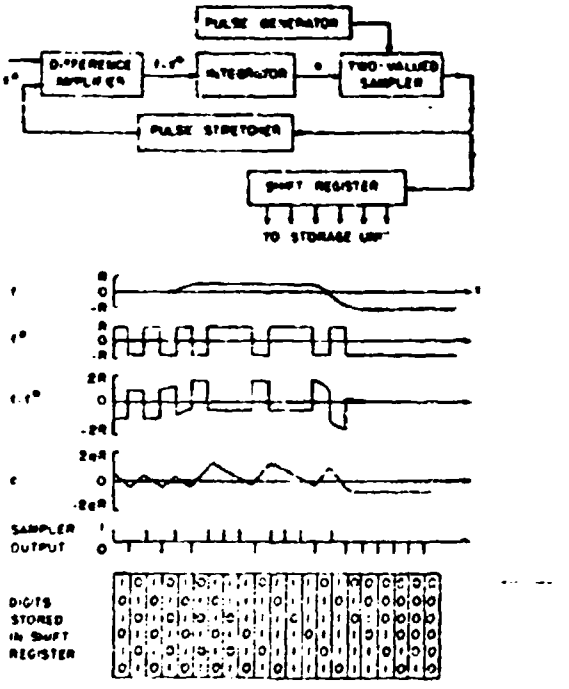

Fig. 2. Quantizer.

digits. Alternatively, we can use a random sequence of digits; the probability is unity that a random sequence will eventually include all possible sequences of length n. A random sequence might be generated by using the quantizer with a white-noise input.

The use of random inputs suggests the application of the apparatus to the synthesis of filters according to statistical criteria (e.g., for an ensemble of inputs consisting of a signal corrupted by noise, in order to minimize the mean-square difference between the signal and the output). The functions u_t (which represent the past of the input) will, in general, form a probability ensemble, and for each u_t the value d(t) of the desired output, of which the output g(t) of the system is to be an estimate, will have a conditional probability distribution. The optimum estimate can be derived by one of many possible criteria from this conditional distribution. In many cases we can approximate the distribution of d(t) conditioned on u_t by the distribution of d(t) conditioned on the quantized approximation, u_t^*, to u_t. We might therefore use a long sample of the input and the desired output, put the input into the quantizer, and for each u^* record samples of the corresponding values of the desired output. If the sample is large enough, we obtain many samples of the desired output for each u^*. We estimate the conditional distribution

from these samples, compute the optimum estimate of the desired output from this distribution, and record the optimum estimate as $\underline{h}(u^0)$. For some estimation criteria there may be easier ways to derive the optimum estimate from the samples; the process described is general.

3.3 EXAMPLE

We now calculate, for a very simple system, the requirements on n and q for synthesis of the system by means of the apparatus described in the preceding section. Consider the linear system with frequency response $a/(s+a)$, that is, with impulse response ae^{-at}. This is a lowpass filter with unity low-frequency gain.

For this system,

$$\underline{h}(u) = \int_0^\infty ae^{-a\tau}\, u(\tau)\, d\tau \tag{20}$$

The system is continuous(R) for every R. Suppose v is in $N_{T,\delta}(u)$. Then whenever $0 \le x \le T$, we have

$$\left| \int_0^x [u(\tau) - v(\tau)]\, d\tau \right| < \delta \tag{21}$$

Then

$$\begin{aligned} |\underline{h}'(u) - \underline{h}(v)|' &= \left| a \int_0^\infty e^{-a\tau} [u(\tau) - v(\tau)]\, d\tau \right| \\ &\le \left| a \int_0^T e^{-a\tau} [u(\tau) - v(\tau)]\, d\tau \right| + a \int_T^\infty e^{-a\tau} [|u(\tau)| + |v(\tau)|]\, d\tau \\ &\le a \left| [e^{-a\tau}(U(\tau) - V(\tau))]_0^T + a^2 \int_0^T e^{-a\tau} [U(\tau) - V(\tau)]\, d\tau \right| + 2R\, e^{-aT} \\ &< a\delta + 2R\, e^{-aT} \end{aligned} \tag{22}$$

in which we have integrated by parts, using

$$U(x) = \int_0^x u(\tau)\, d\tau, \quad V(x) = \int_0^x v(\tau)\, d\tau \tag{23}$$

Next, having selected ϵ as the tolerance on the output, we choose T and δ so that

$$a\delta + 2R\, e^{-aT} \le \epsilon \tag{24}$$

and set $q = \delta/4R$, $n = T/q$, so that

$$4qn + 2\, e^{-4qn} \le \epsilon/R \tag{25}$$

It is difficult to minimize the storage that is required (i.e., to minimize n), but

some bounds are easy to obtain. T must be greater than $(1/a) \log (2R/\epsilon)$, and q must be less than $\epsilon/4aR$; hence, $n = T/q$ must be greater than $(4R/\epsilon) \log (2R/\epsilon)$, which gives a lower bound on n (log = natural logarithm). We obtain an upper bound on the minimum n by finding the n that is required if we arbitrarily choose T and δ so that $a\delta = 2Re^{-aT}$. We obtain $T = (1/a) \log (4R/\epsilon)$, $q = \epsilon/8aR$, and $n = (8R/\epsilon) \log (4R/\epsilon)$.

If we set a tolerance of 10 per cent ($R/\epsilon = 10$), then $120 < n < 300$, approximately, and the minimum number of stored values required is between 10^{36} and 10^{90}.

3.4 CONCLUSIONS

For $R/\epsilon = 10$, a set of 11 numbers can be so determined that the output will always be within ϵ of at least one of them. The requirement of 10^{36} stored values seems, therefore, to imply a very inefficient scheme of synthesis. Part of this large storage requirement arises from the fact that many widely different functions u will have the same $\underline{h}(u)$, and this grouping, since it depends on the system, cannot be built into the apparatus in advance. The same value of $\underline{h}(u)$ therefore must be stored separately for many different functions u. However, there are still several sources of inefficiency that might be corrected.

The inefficiency of the apparatus arises partly from the fact that u_t^* is derived from a quantized approximation to f. This imposes certain constraints on the quantized approximation to u_t. It can be shown that if the quantized approximation u_t^* is derived directly from u_t, so that these constraints are not imposed, we can construct for every u in $PB_{(R,3)}$ a u^* with the property that

$$\left| \int_0^x [u(\tau) - u^*(\tau)]\, d\tau \right| \leq qR \tag{26}$$

instead of $4qR$. Such a u_t^* can be constructed by using n linear filters, of which the k^{th} filter, $k = 1, 2, \ldots, n$, has an impulse response that is 1 from zero to kn and zero thereafter, and by quantizing the output of each filter. Then q can be chosen to be four times as large for the same T and δ, and, in the example of section 3.3, the minimum number of stored values will be between 10^9 and 10^{23}. [Although the quantized output of the k^{th} filter has $k + 1$ possible values, and hence $2^n(n+1)!$ conceivable combinations of values, only 2^n of these combinations are possible if the input is bounded(R).]

There is another source of inefficiency that might be removed. Although data on the past of the input have been taken only from a bounded interval of the recent past, as much data have been taken from the more remote parts of this interval as from the more recent parts, and this is unnecessary. We can correct this by making the q-intervals longer for the remote past than for the recent past. This can be done simply by omitting some of the filters from the set described. This would effect a further substantial

reduction in the number of stored values that is required.

The apparatus that is thus evolved, consisting of linear filters with quantised outputs, is practically similar to, although different in derivation from, the apparatus proposed by Bose (9).

IV. ANALYTIC SYSTEMS

4.1 THEORY

One of the conclusions of Section II was that any system that is continuous(R) can be approximated arbitrarily closely by a polynomial system constructed from linear systems whose impulse responses are Lebesgue integrable. This polynomial system is a sum of a constant, a linear system, and products of linear systems.

The output of a constant system can be represented by a fixed real number; the output of a linear system can be represented by means of a convolution integral. For a product of linear systems, whose impulse response functions are $k_1, k_2, \ldots, k_n$, we have

$$g(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{n} k_i(\tau_i) \prod_{i=1}^{n} f(t - \tau_i)\, d\tau_1 \ldots d\tau_n \tag{27}$$

where f represents the input and g the output. Or, if we write

$$h_n(\tau_1, \ldots, \tau_n) = k_1(\tau_1)\, k_2(\tau_2) \cdots k_n(\tau_n) \tag{28}$$

we have

$$g(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} f(t - \tau_i)\, d\tau_1 \ldots d\tau_n \tag{29}$$

This is, at least in a restricted sense, a system of n^{th} degree, since multiplying the input by a constant A results in multiplying the output by A^n. The function h_n (which is Lebesgue integrable) will be called the system function.

The sum of two n^{th} degree systems is an n^{th} degree system whose system function is the sum of the system functions of the summands. Therefore, the representation of a polynomial system requires only one term of each degree:

$$g(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau)\, f(t-\tau)\, d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1, \tau_2)\, f(t - \tau_1)\, f(t - \tau_2)\, d\tau_1 d\tau_2$$

$$+ \cdots + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_N(\tau_1, \ldots, \tau_N) \prod_{i=1}^{N} f(t - \tau_i)\, d\tau_1 \ldots d\tau_N \tag{30}$$

Any continuous system can be approximated arbitrarily closely in this form.

The domain indicated in these integrals has been chosen as $(-\infty, \infty)$ so that physically unrealizable systems can also be represented. For a physically realizable system, each system function h_n will have the value zero whenever any of its arguments is negative, and the integrals can then be taken on $(0, \infty)$.

For convenience in certain computations, we define the system transforms H_n by

$$H_n(s_1, \ldots, s_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \exp(-s_1\tau_1 - \ldots - s_n\tau_n)\, d\tau_1 \ldots d\tau_n \tag{31}$$

(in particular, $H_o = h_o$). Since the system functions are absolutely integrable, these transforms can be defined as Fourier transforms, with $s_i = j\omega_i$, and as such they will always be well defined. Alternatively, if the system is physically realizable, they can be defined as Laplace transforms, with $s_i \geq 0$. If the system transforms are known, the system functions can be determined (at least theoretically).

The scope of this formula will be extended in two ways. First, we form a power series by allowing an infinity of terms; second, we relax the condition of Lebesgue integrability to a condition of absolute integrability that permits impulses in the system functions. By using impulses, we can approximate some no-memory systems, even though such systems are never continuous(R).

Having extended the scope of the formula, we must impose some restrictions on the system functions in order to guarantee that the formula will specify a unique output for every input in a well-defined set. We define the norms of the system functions by

$$\|h_n\| = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |h_n(\tau_1, \ldots, \tau_n)| \, d\tau_1 \ldots d\tau_n \tag{32}$$

(in particular, $\|h_o\| = |h_o|$). We then require that all these norms be finite and that the power series

$$B_H(x) = \sum_{n=0}^{\infty} \|h_n\| \, x^n \tag{33}$$

have a nonzero radius of convergence R_H. Then, if the input is bounded(R) for any R less than R_H, all the integrals will converge absolutely, the series will converge absolutely, and the output will be bounded($B_H(R)$); that is, the value of the output will never have magnitude greater than $B_H(R)$. The function B_H will be called the bound function of the system $\underline{H}$, and R_H will be called the radius of convergence of the system. A system that can be exactly represented in this form, with these conditions, will be called an analytic system.

The system functions of an analytic system are not unique; in fact, any h_n can be altered by permuting its arguments $\tau_1, \ldots, \tau_n$ without altering the system. However, it can be shown that unique system functions are obtained under the condition that the system functions be symmetric, i.e., that they be unaltered by any permutation of their arguments. If a system is specified with unsymmetric system functions, we can replace each system function by the arithmetic mean of the functions obtained by all possible permutations of its arguments, and the resulting system functions will be symmetric and will define the same system.

It can be shown, by using the concept of multilinear functions, that the analytic system is a generalization of the power series [cf. Hille (14)]. A function f of n arguments is called a symmetric n-linear function if it is symmetric and if it is linear in each argument. We are interested in two cases, the familiar case in which the arguments are

real numbers, and the case in which the arguments are real functions. Any n-linear function of n real numbers $x_1, x_2, \ldots, x_n$ can be written in the form

$$f(x_1, \ldots, x_n) = Ax_1x_2 \ldots x_n \tag{34}$$

where A is a fixed number. The function $\underline{H}_n$, defined by $g = \underline{H}_n(f_1, \ldots, f_n)$, where $g, f_1, \ldots, f_n$ are all real functions and

$$g(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} f_i(t - \tau_i)\, d\tau_1 \ldots d\tau_n \tag{35}$$

where h_n is a real-valued symmetric function of n real variables, is also a symmetric n-linear function. Now define a homogeneous function of n^{th} degree as a function of a single argument that can be derived from a symmetric n-linear function by setting all the arguments equal: $x_1 = x_2 = \ldots = x_n = x$, or $f_1 = f_2 = \ldots = f_n = f$; define a power series as an infinite summation of homogeneous functions. With real-number arguments, we obtain the familiar power series; from the n-linear function with real-function arguments we derive the formula that we have used to represent analytic systems.

4.2 EXAMPLES

Two kinds of examples are to be described here; general analytic systems with particular kinds of inputs, and particular kinds of analytic systems considered without restriction on the input (except boundedness).

In the study of linear systems, particular emphasis is placed on two kinds of inputs: impulses, which place in evidence the physical significance of the impulse response, and sinusoids (or exponentials), which perform a like service for the transform of the impulse response. Sums of impulses, or sums of sinusoids, are trivially accounted for by superposition; but superposition does not hold in nonlinear systems.

Consider an analytic system whose input is a sum of impulses. Strictly, this is not permissible, because a well-defined output is guaranteed only for bounded inputs. However, a formal result can be obtained, and, since laboratory approximations to impulses are really bounded pulses, the formal result has a practical interpretation. For the sake of simplicity, we consider an analytic system with only one term, i.e., all the system functions except one are zero; a general analytic system is a sum of such systems.

Let the input f to such a system be

$$f(t) = \sum_{m=1}^{M} A_m \delta(t - T_m) \tag{36}$$

that is, a sum of M impulses, with the m^{th} impulse having the value A_m and occurring at time T_m. Then the output g of the system is given by

$$g(t) = \sum_{m_1=1}^{M} \cdots \sum_{m_n=1}^{M} A_{m_1} A_{m_2} \cdots A_{m_n} h_n\left(t-T_{m_1}, \ldots, t-T_{m_n}\right) \tag{37}$$

Each permutation of every combination of n impulses (not necessarily n different impulses) chosen from the input thus gives a contribution to the output; this output is equal to the product of the values of these impulses with an appropriate value of the system function. The integrals that appear in the general expression for the output have been, in effect, replaced by sums. If we consider a general input as approximable by a sum of impulses, then the output at any time is a weighted sum of products of past values of the input, the weighting being determined by the system functions. The system functions of an analytic system might therefore be considered as its impulse response functions.

Now suppose the input is a sum of sinusoids, or, since a sinusoid can be expressed as a sum of complex exponentials, suppose the input is a sum of exponentials. Thus

$$f(t) = \sum_{m=1}^{M} A_m \exp(s_m t) \tag{38}$$

Then the output g of the system is given by

$$g(t) = \sum_{m_1=1}^{M} \cdots \sum_{m_n=1}^{M} A_{m_1} A_{m_2} \cdots A_{m_n} H_n\left(s_{m_1}, \ldots, s_{m_n}\right) \exp\left[\left(s_{m_1} + \ldots + s_{m_n}\right)t\right] \tag{39}$$

Each permutation of every combination of n exponentials (not necessarily n different exponentials) chosen from the input gives a contribution to the output that is an exponential with a complex frequency equal to the sum of the frequencies of these exponentials, and with an amplitude that is equal to the product of the amplitudes of these exponentials and an appropriate value of the system transform. Since a sinusoid is a sum of two complex exponentials, each with frequency equal to the negative of the other, these contributions account for the harmonics, sum frequencies, and difference frequencies that we know occur in nonlinear systems. The system transform gives, in terms of the magnitudes and phases of the input sinusoids, the magnitude and phase of each sinusoid in the output. The system transforms might therefore be considered as frequency-response functions.

We now consider some special types of analytic systems. First, we consider the identity system, whose output always equals its input. We shall represent this system by $\underline{1}$. This is a linear no-memory system. All its system functions are zero except the first, i_1, which is a unit impulse; hence, all its system transforms are zero except the first, $I_1(s) = 1$.

Next, we consider two types of systems that are easy to deal with by methods that are already in wide use: linear systems and no-memory systems. A linear system is analytic, with infinite radius of convergence, if its impulse response is absolutely integrable; this impulse response is then its first system function, and all the other system functions are zero. The first system transform is the frequency-response function of the linear system, and all other system transforms are zero. Conversely, an analytic system is linear if all its system functions (or system transforms) are zero except the first.

A no-memory system is analytic if the value of the output is given in terms of the value of the input by a power series, and its radius of convergence is the radius of convergence of this power series. If

$$g(t) = \sum_{n=0}^{\infty} a_n (f(t))^n \tag{40}$$

then the system functions are all products of impulses,

$$h_n(\tau_1, \ldots, \tau_n) = a_n \delta(\tau_1)\, \delta(\tau_2) \ldots \delta(\tau_n) \tag{41}$$

which indicates that the value of the output at any time is independent of past values of the input. The system transforms are all constants,

$$H_n(s_1, \ldots, s_n) = a_n \tag{42}$$

which indicates that if the input is a sum of sinusoids, the amplitude and phase of the output sinusoids are independent of the frequencies of the input sinusoids. Conversely, if the system transforms of an analytic system are all constants, the system is a no-memory system.

4.3 COMBINATIONS OF ANALYTIC SYSTEMS

Because engineering, at the practical level, consists largely of putting things together and making them work, analysis and synthesis have become important parts of the theory of linear systems, and they may be expected to be important in the theory of nonlinear systems as well. Analysis is generally easier than synthesis, and it may be that the best way to develop a good theory of synthesis is to develop first a good theory of analysis. An approach to the analysis of nonlinear systems is proposed in this section and elaborated in subsequent sections.

In the fundamental approach we begin with analytic systems and interconnect them in several ways, with the object of determining when the result of this interconnection constitutes an analytic system, and, whenever it does, what its system functions or system transforms are. Many practical systems can be described as combinations of linear and

no-memory systems, which can be easily represented in analytic form. For such systems, this approach may be, for some purposes, an adequate method of analysis. We begin with some simple forms of interconnection: sums, products, and cascade combinations.

The sum $\underline{H} + \underline{K}$ of two systems $\underline{H}$ and $\underline{K}$ is constructed by tying their inputs together, so that the same input is applied to both, and adding their outputs. Thus, $\underline{Q} = \underline{H} + \underline{K}$ if and only if, for every input f, $\underline{Q}(f) = \underline{H}(f) + \underline{K}(f)$.

If the systems $\underline{H}$ and $\underline{K}$ are analytic, a trivial calculation gives the result that the system functions of $\underline{Q}$ are

$$q_n(\tau_1, \ldots, \tau_n) = h_n(\tau_1, \ldots, \tau_n) + k_n(\tau_1, \ldots, \tau_n) \tag{43}$$

and the system transforms of $\underline{Q}$ are

$$Q_n(s_1, \ldots, s_n) = H_n(s_1, \ldots, s_n) + K_n(s_1, \ldots, s_n) \tag{44}$$

But these results are not sufficient to show that $\underline{Q}$ is analytic, since we still have to show the existence of a bound function and a nonzero radius of convergence. However, this is not difficult. It follows from Eq. 43 that

$$\|q_n\| \leq \|h_n\| + \|k_n\| \tag{45}$$

and from this bound on the norms of the system functions of $\underline{Q}$ we can determine an upper bound on the bound function,

$$B_Q(x) \leq B_H(x) + B_K(x) \tag{46}$$

and a lower bound on the radius of convergence,

$$R_Q \geq \min (R_H, R_K) \tag{47}$$

Note that if $\underline{H}$ is a system of degree N (i.e., all system functions after the N^{th} are zero), and $\underline{K}$ is a system of degree M, then the degree of $\underline{Q}$ will not exceed the larger of the two numbers N, M.

Almost as easily treated is the product $\underline{HK}$ of two systems, constructed by tying the inputs together and multiplying the outputs: $\underline{Q}(f) = \underline{H}(f)\,\underline{K}(f)$. If $\underline{H}$ and $\underline{K}$ are analytic, a straightforward calculation of the output of their product, after terms of like degree are collected, yields

$$q_n(\tau_1, \ldots, \tau_n) = \sum_{i=0}^{n} h_i(\tau_1, \ldots, \tau_i)\, k_{n-i}(\tau_{i+1}, \ldots, \tau_n) \tag{48}$$

and

$$Q_n(s_1,\ldots,s_n) = \sum_{i=0}^{n} H_i(s_1,\ldots,s_i)\, K_{n-i}(s_{i+1},\ldots,s_n) \tag{49}$$

From this, we determine that

$$\|q_n\| \le \sum_{i=0}^{n} \|h_i\| \cdot \|k_{n-i}\| \tag{50}$$

from which it follows that

$$B_Q(x) \ge B_H(x)\, B_K(x) \tag{51}$$

and

$$R_Q \ge \min (R_H, R_K) \tag{52}$$

If $\underline{H}$ is a system of degree N, and $\underline{K}$ is a system of degree M, then the degree of $\underline{Q}$ will not exceed N + M.

The cascade combination $\underline{H} * \underline{K}$ of two systems $\underline{H}$ and $\underline{K}$ is formed by applying the input to $\underline{K}$, using the output of $\underline{K}$ as the input to $\underline{H}$, and taking the output from $\underline{H}$. Then $\underline{Q} = \underline{H} * \underline{K}$ if and only if, for every input f, $\underline{Q}(f) = \underline{H}(\underline{K}(f))$. Note that $\underline{H} * \underline{K}$ and $\underline{K} * \underline{H}$ are not the same, although in special cases (e.g., linear systems) they are equivalent.

The system functions and system transforms of $\underline{Q}$ are given by formulas that are derived, as in the product case, by a straightforward computation of the output, in which terms of like degree are then collected. These formulas are rather complicated in the general case, although they can be expressed fairly simply in certain almost general cases.

The first step in the calculation of q_n or Q_n is to determine, for every positive integer i, all possible permutations of combinations of i non-negative integers whose sum is n. In each permutation these i integers will be called m_j, j=1, 2, ..., i. The system function q_n is given by a convolution-like integral involving h_n and $k_{m_1}, k_{m_2}, \ldots, k_{m_i}$:

$$q_n(\tau_1,\ldots,\tau_n) = \sum_{i=0}^{\infty}\sum \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_i(x_1,\ldots,x_i) \prod_{j=1}^{i} k_{m_j}(\ldots,\tau_{(\ )}-x_j,\ldots)\, dx_1\ldots dx_i \tag{53}$$

The order of the subscript indices on the τ's is immaterial, because permutation of the arguments of a system function does not change the system it characterizes; hence they are not indicated. The only important point to remember is that there should be one of each, from 1 through n, and the τ's might as well be indexed in the order in which they

appear when the k_{m_j} are written in order of increasing j. The second summation sign in this formula indicates a summation over the set of permutations indicated at the beginning of the paragraph.

The system transforms Q_n are given by

$$Q_n(s_1,\ldots,s_n) = \sum_{i=0}^{\infty} \sum H_i(s_1,\ldots,s_i) \prod_{j=1}^{i} K_{m_j}(\ldots,s_{(\ )},\ldots) \tag{54}$$

where $s_{(\ )}$ is the sum of the m_j arguments of K_{m_j}. In this formula, as in formula 53, the second summation is over all permutations of i numbers m_j whose sum is n, and the s-variables may be indexed $1,\ldots,n$ in any order.

In the general case an infinite summation is required for the determination of each system transform or system function. However, when i is greater than n, every combination of m_j must contain at least one m_j that is equal to zero. Therefore, there are two special cases in which each c_n can be determined as a sum of a finite number of terms: when $\underline{H}$ is a system of finite degree, and when $K_o = 0$.

In many practical systems it is possible to obtain a description in which the components are analytic systems with zero constant terms, by first solving for the particular case of zero input to determine the operating points, and then measuring all variables with respect to these operating points. If both systems in a cascade combination have zero constant terms (i.e., $H_o = K_o = 0$), then for the first few system transforms of the cascade combination we obtain

$$Q_o = 0$$

$$Q_1(s) = H_1(s)\,K_1(s)$$

$$Q_2(s_1,s_2) = H_1(s_1+s_2)\,K_2(s_1,s_2) + H_2(s_1,s_2)\,K_1(s_1)\,K_1(s_2) \tag{55}$$

In analytic systems, as in linear systems, the solution of the cascade problem involves integration in the time domain, but does not involve integration if the solution is expressed in the frequency domain.

To determine bounds on norms, bound function, and radius of convergence, in the general case, we obtain

$$\|q_n\| \le \sum_{i=0}^{\infty} \sum \|h_i\| \prod_{j=1}^{i} \|k_{m_j}\| \tag{56}$$

from which it can be shown that

$$B_Q(x) \le B_H(B_K(x)) \tag{57}$$

and

$$R_Q = \min (R_K, \text{ solution of } B_K(x) = R_H) \tag{58}$$

The resultant system $\underline{Q}$ will not always be analytic, there might not be any nonzero radius of convergence if K_o is greater than R_H. However, the same conditions that ensured that the system functions and system transforms of $\underline{Q}$ could be computed by a finite summation will also ensure the analyticity of $\underline{Q}$.

If the components of a cascade combination are of degree N and M, respectively, the degree of the resultant system will not exceed NM.

In all cases, sum, product, and cascade, the bounds on the bound functions are immediately obvious if they are regarded only as bounds on the output. However, the bounds on the bound functions provide more information than this. The coefficients of the power-series expansion of the bound provide upper bounds on the norms of the system functions. Furthermore, the formulas for bounds on the bound functions and radii of convergence remain valid if the bound functions and radii of convergence that appear in the formulas are replaced by bounds that have been obtained from a previous calculation.

The bound functions of combination systems do not have to be computed by power-series methods; they may be computed graphically or numerically. Because the system calculations must be performed by calculating the system functions or transforms one at a time, the bound functions provide a useful method for controlling their accuracy. If, for example, only the first three system functions of a combination system are calculated, a bound on the error can be obtained by subtracting from the upper bound of its bound function the first three terms of its power-series expansion. The resulting function gives, for every R, an upper bound on the magnitude of the error for inputs that are bounded(R).

4.4 EXAMPLES OF CASCADE SYSTEMS

It may often happen in practice that only one system in a cascade chain is nonlinear, and all the others are linear. In such a chain some fairly simple and easily recognizable forms are obtained for the solution.

For a linear system $\underline{L}$ followed by a nonlinear analytic system $\underline{H}$ (i.e., for the combination $\underline{Q} = \underline{H} \circ \underline{L}$), we obtain

$$Q_n(s_1, \ldots, s_n) = H_n(s_1, \ldots, s_n)\, L_1(s_1)\, L_1(s_2) \cdots L_1(s_n) \tag{59}$$

For a nonlinear system followed by a linear system (i.e., for the combination $\underline{Q} = \underline{L} \circ \underline{H}$), we obtain

$$Q_n(s_1, \ldots, s_n) = L_1(s_1 + s_2 + \ldots + s_n)\, H_n(s_1, \ldots, s_n) \tag{60}$$

For a nonlinear system $\underline{H}$ preceded by a linear system $\underline{L}$ and followed by a linear system $\underline{K}$ (i.e., for the combination $\underline{Q} = \underline{K} * \underline{H} * \underline{L}$), we obtain

$$Q_n(s_1, \ldots, s_n) = K_1(s_1 + \ldots + s_n)\, H_n(s_1, \ldots, s_n)\, L_1(s_1) \ldots L_1(s_n) \tag{61}$$

In the particular case in which the nonlinear system $\underline{H}$ is a no-memory system, so that each system transform H_n is a constant A_n, we have

$$Q_n(s_1, \ldots, s_n) = A_n K_1(s_1 + \ldots + s_n)\, L_1(s_1) \ldots L_1(s_n) \tag{62}$$

This form is so easy to recognize that it can be used for synthesis, since any analytic system whose system transforms are of this form must be a cascade combination of linear systems with one no-memory system, or the equivalent of such a combination.

As an illustration of the solution of cascade combination systems, we consider an amplitude-modulated radio communication system. We shall suppose that the carrier and the signal are added in the transmitter, and pass through a nonlinear no-memory device that acts as a modulator. The output of the modulator passes through a radiofrequency amplifier, a propagation path, and several radiofrequency amplifiers in the receiver, all of which are represented by a linear narrow-band filter. Another nonlinear no-memory device acts as a detector, and the output of the detector then passes through a sequence of audio-frequency devices, represented by a linear filter.

The modulator will be assumed to be a second-degree system, whose output y in terms of its input x is given by

$$y = m_1 x + m_2 x^2 \tag{63}$$

The radiofrequency channel will be assumed to be linear with frequency-response function R(s). This filter will be assumed to have zero response at audio frequencies and at the harmonics of the carrier. The detector is another second-degree device, with output y in terms of input x given by

$$y = d_1 x + d_2 x^2 \tag{64}$$

The audio-frequency channel will be assumed to be linear with frequency-response function A(s), and the response of this channel will be assumed to be zero at zero frequency and at radio frequencies.

The fundamental cascade combination formula is then applied three times in succession to obtain the system transforms of the complete channel:

$$Q_0 = 0$$

$$Q_1(-) = m_1 d_1 A(s) R(s)$$

$$Q_2(s_1, s_2) = m_2 d_1 A(s_1 + s_2) R(s_1 + s_2) + m_1^2 d_2 A(s_1 + s_2) R(s_1) R(s_2)$$

$$Q_3(s_1, s_2, s_3) = m_1 m_2 d_2 A(s_1 + s_2 + s_3) R(s_1) R(s_2 + s_3) + m_1 m_2 A(s_1 + s_2 + s_3) R(s_1 + s_2) R(s_3)$$

$$Q_4(s_1, \ldots, s_4) = m_2^2 d_2 A(s_1 + s_2 + s_3 + s_4) R(s_1 + s_2) R(s_3 + s_4) \qquad (65)$$

and all further system transforms are zero. However, Q_1 and the first term of Q_2 are zero, since we have assumed that $A(s) R(s) = 0$, and the two terms of Q_3 can be combined by permuting the variables in one of them. The only nonzero system transforms then become

$$Q_2(s_1, s_2) = m_1^2 d_2 A(s_1 + s_2) R(s_1) R(s_2)$$

$$Q_3(s_1, s_2, s_3) = 2 m_1 m_2 d_2 A(s_1 + s_2 + s_3) R(s_1) R(s_2 + s_3)$$

$$Q_4(s_1, \ldots, s_4) = m_2^2 d_2 A(s_1 + s_2 + s_3 + s_4) R(s_1 + s_2) R(s_3 + s_4) \qquad (66)$$

These transforms characterize the complete channel as a nonlinear analytic system.

Now suppose that the input consists of a number of sinusoids, one at the carrier (radio) frequency, with exponential components at $s = \pm j\omega_r$, and the rest at audio frequencies, $s = \pm j\omega_{a1}$, $\pm j\omega_{a2}$, and so forth. Then no output will be obtained from Q_2, since the only frequencies in the input for which $R(s)$ is nonzero are the positive and negative carrier frequencies, whose sum is either zero or twice the carrier frequency, for both of which $A(s_1 + s_2)$ is zero.

A nonzero output is obtained from Q_3 only when s_1 is a carrier frequency and $s_2 + s_3$ is the sum of a carrier and an audio component, and the two carrier frequencies are of opposite sign; the sum frequency will then be the audio frequency. (For each audio component there will be four terms: s_1 either plus or minus, and with either s_2 or s_3 as the audio component.) Therefore Q_3 gives the demodulated audio output, which is proportional to the audio input, the square of the carrier-frequency input, the audio-frequency gain, the square of the radiofrequency gain, the linear part of the modulator, the second-degree part of the modulator, and the second-degree part of the detector. Q_4 gives a nonzero output only if both $s_1 + s_2$ and $s_3 + s_4$ are sums of audio- and carrier-frequency components, and the sum frequency is the sum of the two audio frequencies. Hence Q_4 gives the harmonic and intermodulation distortion components in the output.

These particular results could also have been obtained by more conventional methods

by assuming the input and then calculating the resulting signals at every point in the system. What we have done here is to solve the system as a system before specifying what the input is to be.

4.5 ANALYTIC FEEDBACK LOOPS

An electrical network consists of a number of elements connected by wires. In many cases, the elements are all two-terminal elements, and each element can be described as a system with one input and one output, by specifying either the voltage in terms of the current or the current in terms of the voltage. The interconnections are expressed in terms of Kirchhoff's laws, which equate to zero either a sum of voltages or a sum of currents. These relations can be expressed by a block diagram or signal-flow graph that contains two kinds of components: systems and summing points. It appears, therefore, that a theory of nonlinear network analysis might be built up by using only two of the three kinds of simple combination described in section 4.3, namely, sums and cascade combinations.

Such a theory has not yet been developed. In terms of the relation between systems and functions, as developed in Section 1, this theory would be essentially a theory of implicit systems. We might hope for an extension to system theory of the fundamental theorem [cf. Rudin (15)] on implicit functions. For the purposes of suggesting the kind of results that such a theory might offer, and of giving a special case with its own useful applications, the solution of additive feedback loops with analytic components is presented in this section. Another special case, the inverse of an analytic system, will be discussed in section 4.7.

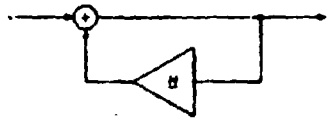

Fig. 3. Simple feedback loop.

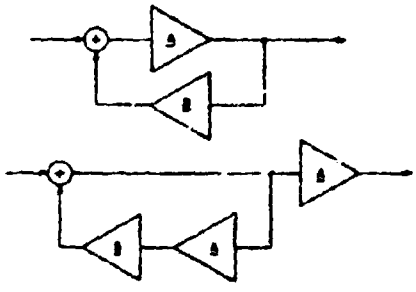

Fig. 4. Reduction of general loop to simple form.

Consider the simple feedback loop illustrated in Fig. 3. This is not such a special case (as it seems) because, as Fig. 4 shows, a general feedback loop can be reduced to two cascade problems and to a feedback problem of the simple form.

Assume that this simple loop system is equivalent to a system $\underline{K}$. Let the input be f, and the output $g = \underline{K}(f)$;

then

$$\underline{K}(f) = f + \underline{H}(\underline{K}(f)) \tag{67}$$

or, as a system equation with the notation of section 4.3,

$$\underline{K} = \underline{I} + \underline{H} \circ \underline{K} \tag{68}$$

where $\underline{I}$ designates the identity system.

Now suppose that $\underline{H}$ is an analytic system. We assume that $\underline{K}$ is analytic, and obtain a formal solution – we shall determine later whether $\underline{K}$ is, in fact, analytic. We shall consider that $H_o = 0$, since a nonzero H_o represents merely a constant added to the input, and we do not have to consider this as part of the feedback loop. Then, for K_o, we obtain

$$K_o = H_1(0) K_o + H_2(0, 0) K_o^2 + H_3(0, 0, 0) K_o^3 + \ldots \tag{69}$$

This equation may have many solutions, but it will always have the solution $K_o = 0$. We assume, as is often the case in practice, that we are looking for the solution that gives zero output for zero input, and hence we accept the solution $K_o = 0$. This allows us to use the simplified form of the cascade equations which occurs when the constant terms of both components are zero.

The next equation that we obtain is

$$K_1(s) = 1 + H_1(s) K_1(s) \tag{70}$$

from which we determine that

$$K_1(s) = \frac{1}{1 - H_1(s)} \tag{71}$$

This is the result that would be obtained from an approximate linear analysis. Next we obtain

$$K_2(s_1, s_2) = H_1(s_1 + s_2) K_2(s_1, s_2) + H_2(s_1, s_2) K_1(s_1) K_1(s_2) \tag{72}$$

and, since K_1 has already been computed, this equation can be solved for K_2, and we have

$$K_2(s_1, s_2) = \frac{H_2(s_1, s_2)}{(1 - H_1(s_1 + s_2))(1 - H_1(s_1))(1 - H_1(s_2))} \tag{73}$$

This procedure can be continued indefinitely, since all the formulas for higher order systems transforms K_n, as derived directly from the cascade formula, will contain K_n only in one term on the right-hand side. Furthermore, since $K_n(s_1, \ldots, s_n)$ appears on the right, multiplied only by $H_1(s_1 + \ldots + s_n)$, the only factors in the denominator

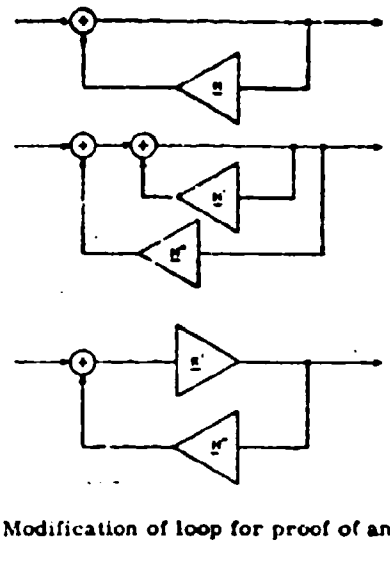

Fig. 5. Modification of loop for proof of analyticity.

of the resulting expression for K_n will be of the form $1 - \Pi_1$, with various arguments.

The form of the solution suggests that the stability of the resultant system $\underline{K}$, and its analyticity as well, might be determined solely from the linear approximation. This is a conclusion that has been reached, in a somewhat different context, by Poincaré and Liapounoff, and it will now be shown that it is essentially true also in this case.

We begin by separating the system $\underline{H}$ into two parts, a linear part $\underline{H}'$, and a part $\underline{H}''$ containing no linear term, as shown in Fig. 5. We then solve the linear part of the feedback loop separately, and obtain a loop with a forward path through the linear system $\underline{K}'$ which is equal to the linear part of $\underline{K}$, and a return path through $\underline{H}''$. The solution of this loop is the same as the solution just obtained; however, we are not interested now in the formal solution for the system transforms, but rather in obtaining bounds on the bound function and the radius of convergence.

For this modified loop we obtain, for an arbitrary input f,

$$\underline{K}(f) = \underline{K}'[f + \underline{H}''(\underline{K}(f))] \tag{74}$$

or, in terms of systems, we have

$$\underline{K} = \underline{K}' \circ (\underline{I} + \underline{H}'' \circ \underline{K}) \tag{75}$$

Since $\underline{K}'$ is linear, by superposition Eq. 74 becomes

$$\underline{K} = \underline{K}' + \underline{K}' \circ \underline{H}'' \circ \underline{K} \tag{76}$$

The known system $\underline{K}' \circ \underline{H}''$ will now be designated by $\underline{Q}$. Then

$$\underline{K} = \underline{K}' + \underline{Q} * \underline{K} \tag{77}$$

Therefore,

$$B_K(x) \leq B_{K'}(x) + B_Q(B_K(x))$$

$$\leq \|k_1\| x + B_Q(B_K(x)) \tag{78}$$

The constant and linear terms of the power-series expansion of B_Q are both zero, so that this equation can be used directly to determine, one at a time, upper bounds on the coefficients of the power-series expansion of B_K: that is, bounds on the norms of the system functions k_n. This procedure is a solution, in power-series form, of

$$y = \|k_1\| x + B_Q(y) \tag{79}$$

and $y(x)$ is an upper bound on $B_K(x)$.

A solution can be obtained without using power series by solving this equation for x in terms of y and graphing the result, as follows

$$x = (y - B_Q(y))/\|k_1\| \tag{80}$$

As y increases from zero, x increases from zero, reaches a maximum, and thereafter decreases, unless the radius of convergence of B_Q (which is equal to R_H) is so small that the curve comes to an end before the maximum value is reached. Equation 80 gives x as an analytic function of y, and investigation of the analytic continuation of this function in the complex plane indicates that its inverse is analytic in a circle about the origin with radius equal to the maximum value of x.

Hence the inverse of this function — which can be obtained immediately from the graph — is the desired upper bound on the bound function B_K. The maximum value of x, which will occur either at the end of the curve (as determined by R_H) or at the turning point, is a lower bound on R_K. Thus the resultant system $\underline{K}$ can be proved analytic if $\|k_1\|$ exists; that is, if the linear approximate solution of the feedback loop has an absolutely integrable impulse response.

Note that the impulse response can be absolutely integrable only if the linear approximation does not include a differential operator and has only damped transients. Stability is therefore a necessary, but not quite sufficient, condition for analyticity.

4.6 EXAMPLE OF FEEDBACK SOLUTION

Section 4.5 constituted essentially a proof of the existence of the solution to certain problems and an outline of the method for their solution. In this section the method is applied to a specific problem, with two changes in the method. First, the feedback loop is solved without first reducing it to the simple form treated in section 4.5. Second,

the result will b proved analytic in spite of the fact that one of its components is not analytic.

The system to be analyzed is the detector circuit of Fig. 6a, which consists of a diode and a capacitor. Since the diode is not perfect, no resistor is needed in parallel with the capacitor. The diode is assumed to have the current-voltage relation

$$i = A(e^{Bv} - 1) \tag{81}$$

and the capacitor has capacitance C.

The relations governing the operation of the circuit can be expressed in the block diagram of Fig. 6b. The system $\underline{D}$ is a no-memory system representing the diode with voltage input and current output; $D_0 = 0$ and

$$D_n(s_1, \ldots, s_n) = AB^n/n! \tag{82}$$

for all other n. The system $\underline{C}$ is linear, representing the capacitor with current input and voltage output; $C_1(s) = 1/Cs$. The system $\underline{C}$ is not analytic.

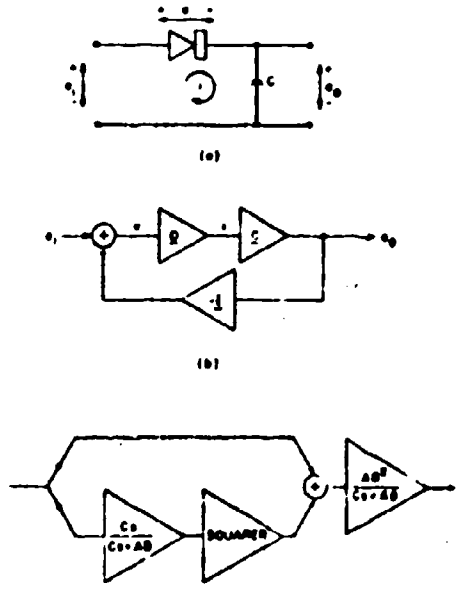

Fig. 6. Detector circuit. (a) Circuit diagram. (b) Block diagram. (c) Synthesis of the approximate solution.

Letting $\underline{H} = \underline{C} \circ \underline{D}$, by a simple application of the cascading formula, we obtain $H_0 = 0$ and

$$H_n(s_1, \ldots, s_n) = \frac{AB^n}{n!C(s_1 + \ldots + s_n)} \tag{83}$$

Now, if we represent the entire detector system by $\underline{K}$, we obtain the system equation from the block diagram.

$$\underline{K} = \underline{H} \circ (\underline{I} - \underline{K}) \tag{84}$$

This is not of the same form as Eq. 68, which was treated in section 4.5, but the same method of solution is applicable. For the first few system transforms, we obtain

$$K_0 = 0$$

$$K_1(s) = \frac{AB}{Cs + AB} \tag{85}$$

$$K_2(s_1, s_2) = \frac{AB^2C^2s_1s_2}{2(C(s_1 + s_2) + AB)(Cs_1 + AB)(Cs_2 + AB)}$$

and further system transforms can be calculated in succession.

The system transform K_2 can be recognized as having the form characteristic of a no-memory system preceded and followed by linear systems, and the second-degree approximation to $\underline{K}$ can thus be synthesized in the form of Fig. 6c from the transforms given in Eqs. 85.

The solution given is only a formal one, in the sense that it will be valid if $\underline{K}$ is analytic, but we do not yet know whether it is analytic. To show analyticity, we view the circuit from a different point of view. We consider the system with input e_i, as before, but with v as the output, and call this system $\underline{P}$. Since $v = e_i - e_o$, we find that $\underline{P} = \underline{I} - \underline{K}$. Therefore, $\underline{K}$ is analytic if and only if $\underline{P}$ is, and these two systems have the same radius of convergence. Furthermore, the norms of their system functions will be the same, except for the first, so that B_P can be used to determine bounds on the error that results from using only a finite number of terms of the expression for $\underline{K}$.

The block diagram for $\underline{P}$ has the simple form of the loop discussed in section 4.5. Since $\underline{C}$ is linear, we find $\underline{H}^e$ and $\underline{H}^o$ by separating out the linear part of $\underline{D}$. For the linear approximation, we obtain

$$P_1(s) = \frac{Cs}{Cs + AB} \tag{86}$$

from which we determine that

$$p_1(\tau) = \delta(\tau) - (AB/C)\, e^{-(AB/C)\tau} \tag{87}$$

and therefore $\|P_t\| = 2$. Proceeding as in the previous section, we find that $Q = -P' * C * D^n$, and when we cascade P' and C we see that the troublesome s in the denominator of C_1 is canceled by the numerator of P_t, so that Q is analytic. From the formula for the upper bound on the bound function of a cascade combination, we find

$$B_Q(x) \leq (e^{Bx} - 1 - Bx)/B \tag{88}$$

The bound $y(x)$ on the bound function $B_p(x)$ is therefore determined by

$$x = y - (e^{By} - 1)/2B \tag{89}$$

The maximum value of x, which is a lower bound on $R_\mu = R_K$, is $(\log 2 - 1/2)/B \approx 0.193/B$. The graphical construction of the upper bound on $B_p(x)$ is shown in Fig. 7.

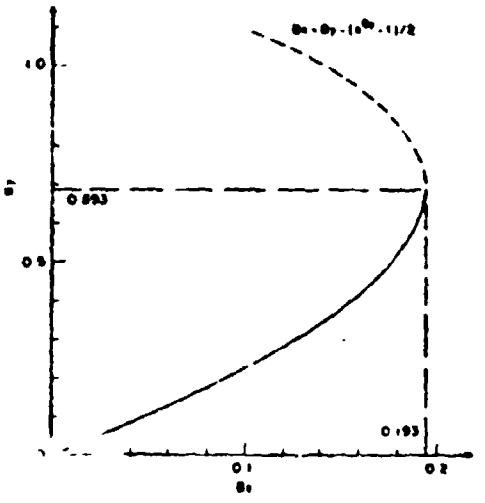

Fig. 7. Bound on the bound function of the detector system.

This result is somewhat disappointing, in that we would expect a much larger, in fact an infinite, radius of convergence. The radius of convergence may in fact be infinite, but the method we have used is inherently incapable of indicating an infinite radius of convergence. It is a conservative method, based essentially on the assumption that the feedback through the nonlinear terms, which we are unable to compute in general except to determine a bound on its magnitude, is exerting the greatest possible effort to make the system unstable. Further research might reveal less conservative tests for analyticity. At any rate, we have proved that the radius of convergence is greater

than zero, and have therefore established some validity for the result of the system calculation.

4.7 THE INVERSE OF AN ANALYTIC SYSTEM

If $\underline{H}$ is a nonlinear system, which produces for every input f in a certain class a corresponding output $g = \underline{H}(f)$, the inverse $\underline{H}^{-1}$ of $\underline{H}$ is a system which, for every input g, produces an output f such that $g = \underline{H}(f)$. Then the cascade combination $\underline{H} * \underline{H}^{-1}$ is the identity system $\underline{I}$.

This does not imply, in general, that $\underline{H}^{-1} * \underline{H} = \underline{I}$, because if $\underline{H}$ can produce the same output for two different inputs, $\underline{H}^{-1}$ cannot tell which of these inputs actually occurred. Generally, we shall say that $\underline{H}$ has a true inverse only if

$$\underline{H} * \underline{H}^{-1} = \underline{H}^{-1} * \underline{H} = \underline{I} \tag{90}$$

The inverse of a physically realizable system is not necessarily physically realizable. This is known in the theory of linear systems; the inverse of a simple delay system would be a simple anticipator.

Within these limitations, however, there are cases in which system inversion is important. For example, a two-terminal network might be designated as a system with current input and voltage output, and we might want to determine its expression with voltage input and current output. Or, we may have a communication channel in which the effect of some system component is to be canceled by introducing another component in cascade.

The problem of determining the inverse of an analytic system is quite similar to the feedback problem. Designating the inverse of $\underline{H}$ by $\underline{K}$, we have the system equation $\underline{H} * \underline{K} = \underline{I}$. As in the feedback problem, an easy solution is possible only if $H_0 = 0$, and if we then choose, out of the many possible solutions for K_0, the solution $K_0 = 0$, which will always exist. We then obtain

$$K_1(s) = 1/H_1(s)$$

$$K_2(s_1, s_2) = \frac{-H_2(s_1, s_2)\, K_1(s_1)\, K_1(s_2)}{H_1(s_1 + s_2)}$$

$$= \frac{-H_2(s_1, s_2)}{H_1(s_1 + s_2)\, H_1(s_1)\, H_1(s_2)} \tag{91}$$

and further system transforms can be calculated successively. It can be verified that these terms also satisfy the equation $\underline{K} * \underline{H} = \underline{I}$, and it may be surmised that this equation is satisfied in general, so that $\underline{K}$ is a true inverse of $\underline{H}$, but a general proof has not been found.

The procedure for proving $\underline{K}$ analytic by determining bounds on $B_K(x)$ and R_K is similar to the procedure for feedback problems. We separate $\underline{H}$ into a linear part $\underline{H}'$

and a part $\underline{H}^0$ with no linear term. Then

$$(\underline{H}^1 + \underline{H}^0) * \underline{K} = \underline{I} \tag{92}$$

The inverse of $\underline{H}^1$ is $\underline{K}^1$, the linear approximation to $\underline{K}$. Cascading each side of Eq. 92 with $\underline{K}^1$, we have

$$\underline{K}^1 * (\underline{H}^1 + \underline{H}^0) * \underline{K} = \underline{K}^1$$

$$\underline{K}^1 * \underline{H}^1 * \underline{K} + \underline{K}^1 * \underline{H}^0 * \underline{K} = \underline{\delta}^1$$

$$\underline{K} = \underline{K}^1 - \underline{K}^1 * \underline{H}^0 * \underline{K} \tag{93}$$

and then proceed as in section 4.5.

We conclude that if the linear approximation to the inverse of an analytic system has an absolutely integrable impulse response, then the inverse is analytic.

The same argument shows also that the inverse is physically realizable if its linear approximation is physically realizable. We have in fact obtained $\underline{K}$ in the form shown in Fig. 5, which, if $\underline{H}$ and $\underline{K}^1$ are physically realizable, describes a physical realization of $\underline{K}$ in terms of physically realizable components.

4.8 MEASUREMENT OF NONLINEAR SYSTEMS

It has been shown that any system that is continuous(R) can be approximated in analytic form, with error that is uniformly smaller than any preassigned tolerance. The proof was purely an existence proof that gave no indication of a method for obtaining an approximation. Although several methods can be used to obtain analytic approximations to given systems, none of these methods can be used in such a way that a given tolerance will be guaranteed.

The problem is similar to that of determining polynomial approximations to a real function. Three methods are available: determination of a polynomial that equals the given function at selected points, as in the method of finite differences; expansion in a Taylor series; and expansion in a series of orthogonal polynomials. All these methods can be generalized and used for systems The first will be described in this section, the second in the section 4.9, and the third will be discussed in Section V.

Consider the time-invariant system $\underline{H}$ as represented by the functional $\underline{h}$, which gives the value of the output at any time in terms of the past of the input. The finite difference of $\underline{h}$ with respect to the real function ϕ is defined as

$$\Delta_\phi \underline{h}(u) = \underline{h}(u + \phi) - \underline{h}(u) \tag{94}$$

This defines $\Delta_\phi \underline{h}$ as a functional, since it specifies a real number for every real function u for which $\underline{h}(u)$ and $\underline{h}(u + \phi)$ exist. We can then consider the system $\Delta_\phi \underline{H}$, which is represented by the functional $\Delta_\phi \underline{h}$, as the finite difference of $\underline{H}$ with respect to ϕ.

Finite differences can be taken successively, with respect to the same or different ϕ-functions:

$$\Delta_{\phi_1\phi_2}\underline{h}(u) = \Delta_{\phi_2}\Delta_{\phi_1}\underline{h}(u) = \Delta_{\phi_1}\underline{h}(u+\phi_2) - \Delta_{\phi_1}\underline{h}(u)$$

$$= \underline{h}(u+\phi_1+\phi_2) - \underline{h}(u+\phi_1) - \underline{h}(u+\phi_2) + \underline{h}(u)$$

$$\Delta_{\phi_1\phi_2\phi_3}\underline{h}(u) = \underline{h}(u+\phi_1+\phi_2+\phi_3)$$

$$- \underline{h}(u+\phi_1+\phi_2) - \underline{h}(u+\phi_1+\phi_3) - \underline{h}(u+\phi_2+\phi_3)$$

$$+ \underline{h}(u+\phi_1) + \underline{h}(u+\phi_2) + \underline{h}(u+\phi_3)$$

$$- \underline{h}(u) \qquad (95)$$

The general form for finite differences of any order can be inferred. These forms can be used for the experimental determination of finite differences.

If $\underline{H}$ is an analytic system, its finite differences will also be analytic. A straightforward calculation, in which terms of like degree are collected, gives the system functions of the first finite difference, if we assume that the system functions h_n are symmetric, for the first few terms:

$$\Delta_\phi h_o = \int_{-\infty}^{\infty} h_1(\tau)\,\phi(\tau)\,d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\,\phi(\tau_1)\,\phi(\tau_2)\,d\tau_1 d\tau_2 + \ldots$$

$$\Delta_\phi h_1(\tau) = 2\int_{-\infty}^{\infty} h_2(\tau,\tau_2)\,\phi(\tau_2)\,d\tau_2$$

$$+ 3\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_3(\tau,\tau_2,\tau_3)\,\phi(\tau_2)\,\phi(\tau_3)\,d\tau_2 d\tau_3 + \ldots$$

$$\Delta_\phi h_2(\tau_1,\tau_2) = 3\int_{-\infty}^{\infty} h_3(\tau_1,\tau_2,\tau_3)\,\phi(\tau_3)\,d\tau_3 + \ldots \qquad (96)$$

Note that h_o does not appear anywhere in these equations, and that, in general, h_n appears only in $\Delta_\phi h_{n-1}$, $\Delta_\phi h_{n-2}$, and so forth. Therefore, if $\underline{H}$ is a system of degree N, the N^{th} finite difference will involve only h_N, and this will appear only in the constant term, so that the N^{th} finite difference is a constant-output system:

$$\Delta_{\phi_1\ldots\phi_N}\underline{h}(u) = N!\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h_N(\tau_1,\ldots,\tau_N)\prod_{i=1}^{N}\phi_i(\tau_i)\,d\tau_1\ldots d\tau_N \qquad (97)$$

Thus an experimental determination of $\Delta_{\phi_1\ldots\phi_N}\underline{h}(0)$ for all possible combinations of real functions $\phi_1,\ldots,\phi_N$ is sufficient to determine the highest order system function h_N. When this has been determined, the undetermined part of the system is a system

of degree N - 1, and therefore all of the system functions can be determined.

In fact, it is not necessary to use combinations of all possible ϕ-functions. If, for example, $\Delta_{\phi_1 \ldots \phi_N} h(0)$ is determined for all possible combinations of functions chosen from some complete normal orthogonal sequence, then we obtain the coefficients of an expansion of h_N in a series of products of these functions.

If we use impulses as the ϕ-functions, designating by δ_x a unit impulse occurring at $\tau = x$, then we have

$$\Delta\delta_{x_1} \ldots \delta_{x_N} h(0) = N!\, h_N(x_1, \ldots, x_N) \tag{98}$$

If we use step functions, designating by $s_x(\tau)$ the value of a function that has value zero for $\tau > x$, and value 1 for $\tau \leq x$, then we have

$$\Delta s_{x_1} \ldots s_{x_N} h(0) = N! \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} h_N(\tau_1, \ldots, \tau_N)\, d\tau_1 \ldots d\tau_N \tag{99}$$

and h_N can be determined by differentiating this.

If we use exponentials, with the notation $e_s(\tau) = e^{-s\tau}$, then

$$\Delta e_{s_1} \ldots e_{s_N} h(0) = N!\, H_N(s_1, \ldots, s_N) \tag{100}$$

An equivalent result can be obtained by using sinusoids instead of exponentials.

If we have a system that is not analytic with finite degree, we can assume that it is approximable by an analytic system of degree N and apply the procedure suggested. We then derive an approximate analytic system that gives the same output as the given system for the inputs that were used to determine the finite differences.

The response of a linear system to any input can be determined if we know its response to a unit impulse, to a unit step, or to all of the sinusoids of a given amplitude. We now have reason to believe that this can be extended to nonlinear systems in the following way. The response of a continuous nonlinear system to any input can be determined, at least approximately, if we know its response to all possible combinations of unit impulses, to all possible combinations of unit steps, or to all possible combinations of sinusoids of all frequencies with a given amplitude.

4.9 TAYLOR-SERIES EXPANSIONS OF ANALYTIC SYSTEMS

Through the theory of functions of a real or a complex variable we have come to associate analyticity with differentiability, as well as with representation in a power series. The differentiation of an analytic system is not only of mathematical interest [cf. Hille (14)], but also useful in the determination of the system functions of an analytic system by a Taylor-series method.

As in vector analysis, for example, when the directional derivative of a scalar function of position depends upon the choice of a direction in space, an analytic system does not have a unique derivative. We shall define the derivative of an analytic system in terms of the functional that represents it, and we shall define it as essentially a directional derivative.

The derivative of the functional $\underline{h}$ with respect to the real function ϕ is defined as the limit of a finite difference quotient:

$$\underline{h}_{\phi}(u) = \lim_{\epsilon \to 0} \frac{\underline{h}(u + \epsilon\phi) - \underline{h}(u)}{\epsilon} \tag{101}$$

This derivative $\underline{h}_{\phi}$ is also a functional. The time-invariant system that it defines will be called $\underline{H}_{\phi}$, the derivative of $\underline{H}$ with respect to ϕ. We can now differentiate $\underline{h}_{\phi}$ with respect to a function ψ, and obtain the second derivative $\underline{h}_{\phi\psi}$, and successive derivatives can be defined ad infinitum.

If $\underline{H}$ is analytic, and its system functions are assumed to be symmetric, then a straightforward calculation of the derivative shows that the derivative is analytic, with system functions

$$h_{\phi,n}(\tau_1, \ldots, \tau_n) = (n+1) \int_{-\infty}^{\infty} h_{n+1}(\tau_1, \ldots, \tau_{n+1})\, \phi(\tau_{n+1})\, d\tau_{n+1} \tag{102}$$

If ϕ is in PBI(M), for any M (that is not necessarily less than R_H), then

$$\|h_{\phi,n}\| \leq (n+1)\, M \|h_{n+1}\| \tag{103}$$

from which it follows that the bound function of $\underline{H}_{\phi}$ is no greater than M times the derivative of the bound function of $\underline{H}$, and its radius of convergence is not less than R_H.

It follows from the assumed symmetry of the system functions h_n that the higher order derivatives with respect to different functions are independent of the order of differentiation. Since, however, a system is not changed by making its system functions symmetric, this conclusion is true whether the derivative is calculated from symmetric or unsymmetric system functions.

Applying the formula (Eq. 102) for the derivative n times, we find that the constant term of the n^{th} derivative is

$$h_{\phi_1 \ldots \phi_n, 0} = n! \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \ldots, \tau_n) \prod_{i=1}^{n} \phi_i(\tau_i)\, d\tau_1 \ldots d\tau_n \tag{104}$$

Thus, h_n can be determined from the values, for $u = 0$, of the n^{th} derivatives of $\underline{H}$. The same comment made in the discussion of finite differences applies here also. The ϕ-functions for which the derivatives must be determined need not be all possible functions, and we can use impulses, step functions, or sinusoids, as in section 4.8. This

result can be used as the basis of a Taylor-series method for the measurement of nonlinear systems. It is not hard to show that

$$h_{\phi_1 \cdots \phi_n, 0} = \lim_{\epsilon \to 0} \Delta_{\epsilon\phi_1 \cdots \epsilon\phi_n} h(0)/\epsilon^n \tag{105}$$

so that, for the application of the Taylor-series method, we can calculate the higher order differences as limits of finite-difference quotients, computed for $u = 0$, instead of by computing successive derivatives for all u.

The expression for the system functions in terms of derivatives provides a proof of the statement that the symmetric forms of the system functions are unique.

The Taylor-series method can be used only for analytic systems, since it derives the system functions from the small-signal response of the system, and therefore determines the large-signal response from the small-signal response by extrapolation. Its advantage over the method of finite differences is that it is not restricted to systems of finite degree. As an experimental method, it has the disadvantage that it involves limits of observed values, so that many observations may be necessary to obtain each limit, while the method of finite differences requires only the observation of output values for specific inputs.

4.10 SYNTHESIS OF ANALYTIC SYSTEMS

Two methods will be mentioned for the synthesis of systems from their analytic representations. The first, obviously applicable only in special cases, consists in recognizing already familiar forms in the system functions. It is illustrated in section 4.6, in which the second-degree approximation to the solution of a feedback problem was synthesized in cascade form. Continued investigation of the properties and applications of analytic representations can be expected to make this method applicable to an ever-widening class of systems.

The second method is based on the fact that if

$$h_n(\tau_1, \ldots, \tau_n) = \prod_{i=1}^{n} \phi_i(\tau_i) \tag{106}$$

then this term of the system can be synthesized as a product of linear systems; if h_n is a sum of such products, then it can be synthesized as a sum of products of linear systems. Such a product expression can be obtained by expanding h_n in a series of products of orthogonal functions. However, no investigation has been made for determining the conditions under which this method can be expected to yield arbitrarily small error.

Neither these methods nor the methods proposed by Wiener and Bose can be considered satisfactory methods of synthesis. All of these methods are capable of the following absurdity: we are given a black box, we examine it in the laboratory by applying inputs and observing the outputs, we assemble several hundred dollars' worth of equipment to approximate its characteristics, and then open the box and find that it contains a handful

of components, the total cost of which is one dollar. An example of this is found in the detector circuit analyzed in section 4.6. By no known method of synthesis of nonlinear systems can we derive, from any expression of the output of this detector in terms of its input, a synthesis in the form in which the circuit was originally given.

A method of synthesis that meets the criteria implied by the foregoing discussion is a practical necessity. The discovery of such a method is one of the great unsolved problems of the theory of nonlinear systems.

V. ORTHOGONAL SYSTEMS

5.1 MEASURE AND INTEGRAL

The purpose of this section is to show, in a general way, how the expansion of real functions in series of orthogonal functions can be extended to indicate a method for expanding nonlinear systems in series of orthogonal systems. Since the expansion of real functions by this method involves the process of computing the integral of a real function, the application of the method to systems will require integration of systems, or at least integration of functionals. The idea of integration must therefore be extended from real functions to systems or functionals.

For this purpose it is appropriate to indicate here some of the fundamental ideas of integration. In this section the relation between integration and measure theory will be roughly outlined, and it will be shown that the theory of probability plays an important part in the integration of functionals.

We begin with the integration of real functions. Suppose we have a real function f, and want to integrate it over the interval (A, B). The integral that we obtain is defined in geometric terms as the net area between the x-axis, the graph of $y = f(x)$, and the lines $x = A$, $x = B$. This area is obtained, in principle, as the limit of a sequence of approximations, each approximation being obtained by dividing the area into a number of strips and estimating the width and the length of each strip. In general, the strips are parallel to the y-axis.

In the classical definition of the Riemann integral the strips are constructed by dividing the interval (A, B) into a large number of small subintervals and taking each subinterval as defining the width of a strip. The length of each strip is estimated as some value assumed by $f(x)$ on the subinterval that defines its width. The integral is estimated as the sum of the areas of the strips, and the integral is defined as the limit of these estimates as the interval (A, B) is divided into smaller and smaller subintervals. If f is continuous, the range over which $f(x)$ varies on each subinterval becomes arbitrarily small as the subintervals are made small, and the estimate of the length of the strip becomes better and better, so that the limit that defines the integral exists. But if f is not continuous, $f(x)$ may continue to vary widely no matter how small the subinterval is made, and the integral thus defined will not exist.

To get around this difficulty, we redefine the strips in such a way that $f(x)$ can not vary widely. We divide the y-axis, instead of the x-axis, into subintervals, so that the length of each strip can be estimated with error that is less than the width of the subinterval, and define the width of the strip as the total length of the set of all x for which $f(x)$ is in the subinterval. If f is continuous, this set of x's will consist of a collection of discrete intervals, and the total length of this set is simply the sum of the lengths of these intervals. But if f is not continuous, this set may not be a collection of intervals, and we are faced with the problem of defining the total length of a set of numbers that is not an interval and cannot be decomposed into intervals.

The Lebesgue theory of measure, by providing at least a partial solution to this problem, achieves the desired generalization of integration. This theory shows how we can assign, to each of a large collection of sets called measurable sets, a number called its measure, so that the measure of a set has the properties that we associate with the idea of total length: the measure of a set is always non-negative, and the measure of the union of a finite or countable collection of nonoverlapping sets is the sum of the measures of the component sets. In particular, the measure called Lebesgue measure has the property that the Lebesgue measure of an interval is the same as its length. By means of Lebesgue measure we can proceed to define integration for many functions that are not continuous: for each subinterval of the range, we multiply the measure of the corresponding subset of the domain by some value in the subinterval of the range, take the sum of these products as an estimate of the integral, and define the limit of these estimates, as the subintervals are made smaller, as the Lebesgue integral of the function.

The point in which we are interested here is not that the Lebesgue integral is defined for functions whose Riemann integral is not defined, but rather that the theory of integration in terms of measure can be used to define integrals of functions that are not real functions. Consider any real-valued function, whose domain may be a set of any kind of objects. Divide the range of the function into subintervals, and for each subinterval consider the subset of the domain on which the value of the function lies in that subinterval. If, to every such subset of the domain, we can assign a number that we can call its measure, then we can proceed, just as in the case of the Lebesgue integral, to define the integral of the function.

The problem now is to define a measure on a set of real functions, and the solution comes from the theory of probability. A probability ensemble is a set of objects, in which we assign to every subset – or at least to certain subsets – a number called the probability of that subset. The probability of the whole ensemble is unity; the probability of every subset is non-negative, and does not exceed unity, and, in fact, probability has all the properties that are required of a measure.

Therefore, if we have a real-valued function whose domain is a probability ensemble, we can define an integral of that function. We divide the range of the function into subintervals, and for each subinterval we multiply some value in that subinterval by the probability that the value of the function will lie in that subinterval. We add these products over all subintervals to obtain an estimate of the integral. It will be seen that by this process we have obtained an estimate of the ensemble average, or expectation, of the value of the function. An integral, defined in terms of probability measure, is simply an ensemble average.

To define the integral of a functional, all we need is an ensemble of real functions u. The integral of the functional is then the ensemble average of the value of the functional for this ensemble of u. The integral of a time-invariant system can then be defined if we have a stationary ensemble of inputs f. The stationariness of the ensemble implies that the ensemble of functions u_t, $u_t(\tau) = f(t-\tau)$, will be the same for every t, and the

integral of the system is defined as the integral, for this ensemble of u_t of the functional that represents it; that is, the ensemble average of the value of the output of the system. In many practical situations, the stationary ensemble of inputs is prescribed by the application for which the system is being considered.

5.2 EXPANSIONS IN ORTHOGONAL SYSTEMS

Every time-invariant system can be represented by a functional, and every stationary ensemble of functions f corresponds to some ensemble of functions u_t which is the same as the ensemble of functions u_t for any t. Therefore, we shall do all our mathematical work in terms of functionals; and the results can be translated immediately in terms of systems by the fact that the average of the value of a functional will equal the average of the value of the output of the system that it represents.

Suppose we have available in the laboratory a bank of nonlinear time-invariant systems $\underline{Q}_i$, and suppose we also have an unknown time-invariant system $\underline{H}$ that is to be approximated as a linear combination of the systems $\underline{Q}_i$. Representing these as functionals, we shall determine real numbers c_i with the property that the functional

$$\underline{h}_0 = \sum_{i=1}^{N} c_i \underline{q}_i \tag{107}$$

is an approximation to $\underline{h}$. The error of approximation will be the output of the system $\underline{E} = \underline{H} - \underline{H}_0$, represented by the functional

$$\underline{e} = \underline{h} - \underline{h}_0 = \underline{h} - \sum_{i=1}^{N} c_i \underline{q}_i \tag{108}$$

We shall obtain an approximate representation of $\underline{H}$ in terms of the systems $\underline{Q}_i$. This representation may be useful in two distinct ways. If the systems $\underline{Q}_i$ are easy to construct, then we have a way of constructing $\underline{H}$ or an approximation to it. If the systems $\underline{Q}_i$ have convenient mathematical representations, we obtain a convenient mathematical representation of $\underline{H}$.

Suppose that the criterion of approximation is that the mean-square value of the error, for a particular stationary ensemble of inputs, be as small as possible. We shall designate the mean or expectation of the value of a functional by an integral sign, to show that our mathematics is analogous to the mathematics of real functions. Then we must determine the numbers c_i so that

$$\int \underline{e}^2 = \int (\underline{h} - \underline{h}_0)^2$$

$$= \int \underline{h}^2 - 2 \sum_{i=1}^{N} c_i \int \underline{h}\underline{q}_i + \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j \iint \underline{q}_i \underline{q}_j \tag{109}$$

as a minimum. At the minimum value, the partial derivatives

$$\frac{\partial}{\partial c_k}\int \varepsilon^2 = -2\int b g_k + 2\sum_{i=1}^{N} c_i \int g_i g_k$$

$$= -2\int \varepsilon g_k \qquad (110)$$

must all be zero. The condition for minimum mean-square error is therefore that, for every k,

$$\int \varepsilon g_k = 0 \qquad (111)$$

It is possible to implement this scheme in the laboratory. The output of each of the systems $\underline{G}_i$ is passed through an adjustable-gain amplifier, which provides for the adjustment of c_i, and the outputs of these amplifiers are added to construct the system $\underline{H}_e$. The output of $\underline{H}_e$ is subtracted from the output of $\underline{H}$ to obtain the error system $\underline{E}$. Now we must multiply the output of each of the systems $\underline{G}_i$ by the output of $\underline{E}$ and obtain the ensemble average of this product. If the ensemble of inputs is ergodic, this ensemble average will equal a time average, and (as Wiener has suggested) we might combine the multiplication with the estimation of the time average by using an overdamped electrodynamometer. Then we have a bank of N meters, and we must adjust the N gain controls so that each meter is made to read zero.

In the general case this may be a troublesome procedure, because the adjustment of a single gain control may change the readings of all the meters. However, if the systems $\underline{G}_i$ are such that

$$\int g_i g_k = \begin{cases} 1, & \text{if } i = k \\ 0, & \text{if } i \neq k \end{cases} \qquad (112)$$

then

$$\int \varepsilon g_k = \int b g_k - c_k \qquad (113)$$

so that the reading of the k^{th} meter is affected only by the k^{th} gain control, and the appropriate adjustment can be made quite simply. Furthermore, if the meters and the gain controls are appropriately calibrated, we can perform the adjustment by setting each gain control to zero, reading each meter, and then setting

$$c_i = \int b g_i \qquad (114)$$

which will give the desired approximation immediately.

The condition imposed on the systems $\mathcal{Q}_i$ can be expressed in a terminology that is conventional for the analogous situation in the theory of real functions. We shall say that the systems $\mathcal{Q}_i$ are all normalized, and that all are orthogonal to each other. The set of systems will be called an orthonormal, or normal orthogonal, set. The approximation obtained by this procedure will be called an expansion in orthogonal systems. Since these conditions were imposed in connection with a particular ensemble of inputs, we shall speak of systems normal and orthogonal with respect to a particular input ensemble.

If the systems $\mathcal{Q}_i$, as given, are not normal and orthogonal with respect to the particular input ensemble that we intend to use, we can construct, by means of a well-known procedure, a set of linear combinations of them that are normal and orthogonal. This procedure can be described by supposing that we have already constructed a set of n normal and orthogonal systems, and we have a system $\mathcal{Q}_{n+1}$ that is not normal or orthogonal to these n systems. We are to construct a linear combination of these $n + 1$ systems which is normal and orthogonal to the first n. We do this by constructing, with the use of the first n systems, a minimum-mean-square-error approximation to $\mathcal{Q}_{n+1}$. The system whose output is the error of this approximation is orthogonal to the first n systems, and if it is not equivalent to zero, then we can normalize it by multiplying it by an appropriate constant. (If it is equivalent to zero, then every linear combination of the $n + 1$ systems can be also obtained with the first n, and the additional system $\mathcal{Q}_{n+1}$ is of no use to us.)

In general, then, all we need to obtain minimum-mean-square-error approximations is a bank of nonlinear systems, some adjustable-gain amplifiers, and some product-average meters. The given systems can be orthogonalized and can then be used to obtain orthogonal expansions of any given system.

How close can these approximations be made? Can the mean-square error of approximation be made arbitrarily small by using a large enough bank of nonlinear systems $\mathcal{Q}_i$? For systems that are continuous(R), considered only for ensembles of inputs that are bounded(R), we have a ready answer. We begin with a sequence of linear systems $\mathcal{K}_i$, with impulse response functions k_i, such that there is no function u not equivalent to zero for which

$$\int_0^\infty u(\tau)\, k_i(\tau)\, d\tau = 0 \qquad (115)$$

for all i. (The sequence of systems whose impulse responses are the Laguerre functions has this property.) Then we form a sequence of nonlinear systems consisting of a constant-output system, these linear systems, and all products of combinations of these systems. The set of all linear combinations of these nonlinear systems is an algebra that separates points (cf. section 2.3), and therefore any system that is continuous(R) can be approximated arbitrarily closely by such linear combinations, in the sense that for any positive number ϵ there exists a linear combination of these systems

whose output never differs from the output of the given system by more than ϵ. It follows immediately that the mean-square error can thus be made less than ϵ^2. The method of orthogonal expansions will yield approximations with mean-square error that is as small as can be obtained, so that, by using a sufficiently large number of systems, the mean-square error can be made arbitrarily small.

This does not imply that the method of orthogonal expansion can be made to yield approximations with uniformly small error. However, it is easily seen that the probability of an error of magnitude greater than any number A cannot exceed $(\epsilon/A)^2$ if the mean-square error is ϵ^2 or smaller, so that we have (for the ensemble with respect to which the expansion was made) an almost uniformly small error in a statistical sense.

An expansion made with respect to one input ensemble will have small error, although not minimum error, when some other input ensembles are used. Consider two ensembles of inputs, E_1 and E_2, consisting of the same set of inputs with different probability distributions. Suppose that the system $\underline{H}$ has been approximated by means of an orthogonal expansion with respect to E_1, so that (if we attach the ensemble designation to the integral sign) we have

$$\int_{E_1} \underline{e}^2 < \delta \tag{116}$$

If the probability of any subset in ensemble E_2 is never greater than M times the probability of the same subset in E_1, then obviously

$$\int_{E_2} \underline{e}^2 < M\delta \tag{117}$$

Then by making the mean-square error that is measured with E_1 sufficiently small, we can make the mean-square error that is measured with E_2 as small as may be required. Thus a small mean-square error in one ensemble implies a small mean-square error in another ensemble. Furthermore, even under the looser condition that every set with zero probability in E_1 must have zero probability in E_2, we can conclude that with input ensemble E_2 there must be small probability of large error, even though no bound on the mean-square error can be set. There is reason to believe that, if the system that is to be approximated is continuous, even this condition can be relaxed, and that there may be orthogonal methods of deriving approximations with uniformly small error. (It is known that such methods exist for the approximation of real functions.)

Some of the preceding discussion is applicable to the expansion, in orthogonal systems, of systems that are not continuous (for example, hysteretic systems). In this case, if the nonlinear systems $\mathcal{G}_i$ are continuous (as they will be if they are products of linear systems with Laguerre-function impulse responses), an approximation with uniformly small error cannot be obtained. However, it may be possible to obtain

approximations with small mean-square error, perhaps even arbitrarily small. At any rate, the smallest mean-square error that is possible with a continuous approximation can be approached.

5.3 FILTERING AND PREDICTION

The same method can also be used in minimum-mean-square-error filtering and prediction problems, if the joint ensemble of input and desired output can be produced in the laboratory. We proceed precisely as though the desired output had been obtained as the output of a system to which the input was applied. The results of the preceding section then imply that, with any bank of nonlinear filters Q_j that is adequate to approximate a continuous system with arbitrarily small mean-square error, we can come arbitrarily close to any performance that can be achieved with a continuous filter.

Note that there may be, in some problems, no optimum continuous filter, if the optimum filter is not realizable. (A heuristic illustration is found in the theory of real functions: the [illegible] extension to a square wave.

References

1. N. Minorsky, Introduction to Nonlinear Mechanics (J. W. Edwards Press, Ann Arbor, Michigan, 1947).
2. V. Volterra, Leçons sur les Fonctions de Lignes (Gauthier-Villars, Paris, 1913).
3. N. Wiener, Differential-Space, J. Math. Phys. 2, 131-174 (1923).
4. N. Wiener, Response of a non-linear device to noise, Report 129, Radiation Laboratory, M.I.T., April 6, 1942.
5. S. Dehara, A method of Wiener in a nonlinear circuit, Technical Report 317, Research Laboratory of Electronics, M.I.T., Dec. 10, 1951.
6. R. Deutsch, On a method of Wiener for noise through nonlinear devices, IRE Convention Record, Part 4, 1955, pp. 186-192.
7. R. C. Booton, Jr., The measurement and representation of nonlinear systems, Trans. IRE, vol. CT-1, no. 4, pp. 32-34 (1954).
8. A. G. Bose, The Wiener theory of nonlinear systems, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Oct. 15, 1954, p. 55.
9. A. G. Bose, A theory of nonlinear systems, Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956.
10. H. E. Singleton, Theory of nonlinear transducers, Technical Report 160, Research Laboratory of Electronics, M.I.T., Aug. 12, 1950.
11. L. A. Zadeh, A contribution to the theory of nonlinear systems, J. Franklin Inst. 255, 387-408 (1953).
12. W. Rudin, Principles of Mathematical Analysis (McGraw-Hill Publishing Company, New York, 1953).
13. E. Hille, Functional Analysis and Semi-Groups, Colloquium Publications, Vol. 31 (American Mathematical Society, New York, 1948).
14. Ibid., p. 65 ff.
15. W. Rudin, op. cit., p. 181.