CONTINUOUS NONLINEAR SYSTEMS

DONALD A. GEORGE

TECHNICAL REPORT 355

JULY 24, 1959

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 355

July 24, 1959

CONTINUOUS NONLINEAR SYSTEMS

Donald A. George

This report is based on a thesis submitted to the Department of Electrical Engineering, M. I. T., July 24, 1959 in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

A functional representation, which is a generalization of the linear convolution integral, is used to describe continuous nonlinear systems. Emphasis is placed on nonlinear systems composed of linear subsystems with memory, and nonlinear no-memory subsystems. An "Algebra of Systems" is developed to facilitate the description of such combined systems. From this algebraic description, multidimensional system transforms are obtained. These transforms specify the system in much the same manner as one-dimensional transforms specify linear systems. The system transforms and the transform of the system's input signal are then used to determine the transform of the output signal. Transform theory is also used for determining averages and spectra of the system output when the input is a random signal Gaussianly distributed. Certain theoretical aspects of the functional representation are discussed.

TABLE OF CONTENTS

CONTENTS

I. INTRODUCTION

1.1 SYSTEM ANALYSIS

In physical analysis, a "system" is often used to specify the relation between a cause and an effect. In system terminology, the cause is the system input and the effect is the output. This is represented in Fig. 1, where x is the input signal and f is the output signal. Usually these signals are functions of time. Of the several general classifications of systems, the class that has been most successfully studied is the linear, time-stationary system. This report is concerned with the nonlinear, stationary system – particularly the continuous nonlinear stationary system. The continuous nonlinear system will be described in detail in section 6.8. The continuous concept implies a certain degree of smoothness in the system's input-output relation. The linear system can be regarded as a special case of the continuous nonlinear system.

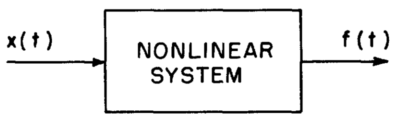


Fig. 1. Nonlinear system.

The analysis of a system is dependent upon finding a mathematical description of the relation between the system input and the system output. Classically, the relation is obtained by means of a differential equation. However, the present means of representing a linear system is by the convolution integral and its associated transforms. The mathematical representation for nonlinear systems which forms the basis of this report is closely related to these modern methods for linear system analysis.

1.2 FUNCTIONAL REPRESENTATION

A function f operates on a set of variables x to produce a new set of variables f(x). A functional, however, operates on a set of functions and produces a new set of functions. In other words, a functional is a function of a function.

The mathematical description used in this report to represent a nonlinear system is the functional series:

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

$$+ \dots + \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_n(\tau_1, \dots, \tau_n)\, x(t-\tau_1) \dots x(t-\tau_n)\, d\tau_1 \dots d\tau_n + \dots \qquad (1)$$

where f(t) is the system output, and x(t) is the system input. The first term in this series is the ordinary convolution integral that is used for linear system analysis. The other terms are generalizations of this linear convolution term. In linear system theory, the function $h_1(t)$ is known as the "impulse response." In section 1.5, the function $h_n(t_1, \dots, t_n)$ will be shown to be a generalized impulse response. In this report, the limits of integration, unless otherwise indicated, will be from $-\infty$ to ∞.

1.3 HISTORICAL NOTE

These functionals were studied by Volterra (1) early in the twentieth century. In 1942, Wiener (2) first applied the functional series to the study of a nonlinear electrical circuit problem. He was concerned with computing the output moments of a detector circuit with a random input. Later, he used this representation as the basis for a canonical form for nonlinear systems (3).

More recently, the functional representation has been investigated by a number of workers. Bose (3) investigated the canonical form problem and developed a system that overcame many of the difficulties associated with Wiener's system. Brilliant (4) was concerned with the validity of the functional representation, and found that systems satisfying a certain continuity condition could be represented. He also showed that the representation was well suited to the combining of nonlinear systems.

Wiener and others have extended the application of this functional representation, in the random input case, beyond the results of Wiener's paper of 1942. Wiener (5) developed the rigorous theory for random (white Gaussian) inputs and applied the theory to such situations as are found in FM spectra. Barrett's (6) paper is an excellent exposition of the state of this theory at the time the present work was undertaken.

1.4 COMPARISON OF THE FUNCTIONAL APPROACH WITH OTHER NONLINEAR METHODS

The analysis of nonlinear systems has been an interesting problem for many years. It is therefore of benefit to compare the present state of the functional approach with the principal classical methods. There are two main classes of solutions to nonlinear problems: transient solutions, and steady-state solutions.

Transient solutions are obtained classically by the solution of nonlinear differential equations (7). For first-order equations, solutions can usually be obtained formally, although numerical integration procedures may be required. However, only special forms of second-order equations can be solved. Force-free solutions for second-order equations can be found with the phase-plane method — even for extremely violent nonlinearities. Examples of violent and nonviolent nonlinearites are shown in Fig. 2. Generally, numerical techniques must be used to solve higher-order equations.

Sinusoidal steady-state solutions can be obtained for systems in which the first harmonic is the only significant term. This is the basis for the "describing function method" (8), and for some others. System order is not a limiting factor, nor, generally, is the violence of the nonlinearity.

The functional series (Eq. 1) is a very general method for representing nonlinear systems (see secs. 2.1 and 6.8). However, at least in the present state of these methods, it does have a definite practical limitation. If the nonlinearities in a system are too violent, the number of terms required for a close approximation becomes very large.

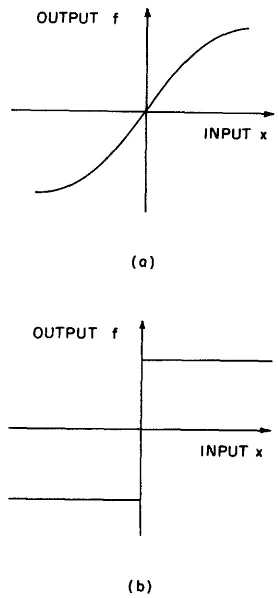


Fig. 2. Violence of nonlinearities: (a) nonviolent nonlinearity (vacuum tube with "medium" signal); (b) violent nonlinearity (ideal clipper).

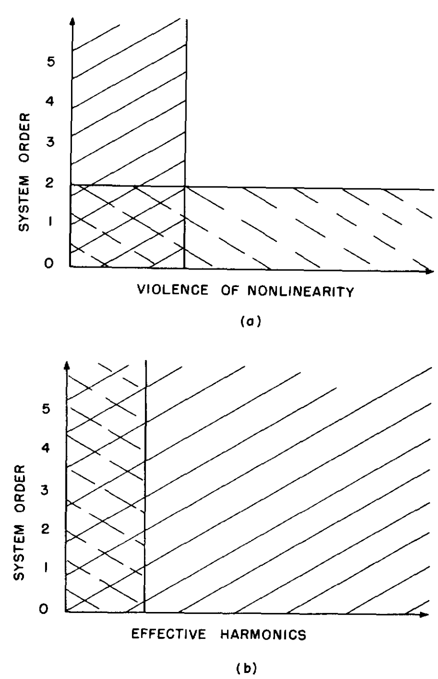


Fig. 3. Application graphs: (a) transient solutions; (b) steady-state solutions. (Unbroken lines show the region covered by the functional representation; broken lines show the region covered by the classical methods.)

It would then be necessary to resort to a computer, and a great deal of the value of the method would be lost. However, if the nonlinearities are sufficiently smooth, the transient response of a system is determined by the first few terms of the series, and there is little limitation from system order. Also, steady-state solutions do not require that the first harmonic be the only significant term.

The comparisons that have just been made are illustrated graphically in Fig. 3. The shaded areas show the regions of effectiveness of various methods of analysis. However, the graphs should not be taken to mean that these methods can cover all systems in the shaded regions, but only a significantly large number.

The first problem in system analysis is to find a suitable mathematical description. This description is called the "system representation." The functional representation studied in this report has three important properties:

(a) It has an explicit input-output relation.

(b) It facilitates the combination of systems.

(c) It allows the consideration of random inputs.

If a representation has an implicit – rather than explicit – input-output relation, it means that the whole problem must be re-solved for each different input. (The differential equation representation is implicit.) Property (b) is important because the electrical engineer spends a great deal of time "putting things together." The effect of random inputs is a problem of great interest to the engineer.

The classical methods based on the differential equation have none of these properties. On the other hand, the significance of transform and convolution methods in linear system analysis rests heavily on these properties. Therefore, these three properties give three distinct advantages to the functional representation as compared with the classical nonlinear methods.

1.5 INTERPRETATION OF THE FUNCTIONAL SERIES

Having indicated the position of the functional representation in the general field of nonlinear system analysis, the writer would like to present an interpretation of the functional series. First of all, the series (Eq. 1) can be viewed as a series of time functions:

$$f(t) = f_1(t) + f_2(t) + \dots + f_n(t) + \dots \tag{2}$$

where

$$f_n(t) = \int \dots \int h_n(\tau_1, \dots, \tau_n)\, x(t-\tau_1) \dots x(t-\tau_n)\, d\tau_1 \dots d\tau_n \tag{3}$$

That is, at some time t_1 we have a series of numbers $f_n(t_1)$ that add up to give the actual value of the system output $f(t_1)$. Also, each of the functions $f_n(t)$ is seen from Eq. 3 to be the result of a convolution operation on the input time function $x(t)$. The first term, $f_1(t)$, in particular, is recognized as being the result of putting $x(t)$ into a linear system with an impulse response, $h_1(t)$. Indeed, each term $f_n(t)$ can be viewed as the output of a system with input $x(t)$.

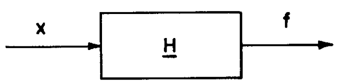


Fig. 4. Nonlinear system.

To take advantage of this idea, we introduce an operator notation. In this notation, if we have a general nonlinear system with output $f(t)$ and input $x(t)$, as illustrated in Fig. 1, then we can write $f(t) = \underline{H}[x(t)]$, or if we make the time dependence implicit, $f = \underline{H}[x]$. Then the symbol $\underline{H}$ represents the operation that the system makes on input x to produce output f. In diagram form, a nonlinear system is then represented as shown in Fig. 4. (The usual operator notation $\mathscr{H}$ is replaced by $\underline{H}$ in this report.)

The first term can be viewed as a linear system operation, and therefore

$$f_1(t) = \underline{H}_1[x(t)]$$

or

$$f_1 = \underline{H}_1[x]$$

where the subscript "1" is added to the $\underline{H}$ notation to denote that the operation is linear. Now, a linear system is specified by means of its impulse response; and thus, associated with the linear system $\underline{H}_1$, there is an impulse response $h_1(t)$, and

$$f_1(t) = \int h_1(\tau)\, x(t-\tau)\, d\tau$$

Now, the second term in the series (Eq. 2) is

$$f_2(t) = \int\int h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \tag{4}$$

If the input $x(t)$ is changed by a gain factor ϵ to give a new input $\epsilon x(t)$, then the new output, $g_2(t)$, is

$$g_2(t) = \int\int h_2(\tau_1, \tau_2)\, \epsilon x(t-\tau_1)\, \epsilon x(t-\tau_2)\, d\tau_1 d\tau_2$$

or

$$g_2(t) = \epsilon^2 f_2(t)$$

Thus, the second term is the result of a quadratic (or squaring) operation. In the operator notation, then, $f_2 = \underline{H}_2[x]$, where the subscript "2" indicates that this is a quadratic operation. Similarly, $f_3 = \underline{H}_3[x]$ and, in general, $f_n = \underline{H}_n[x]$. Associated with each $\underline{H}_n$ is the function $h_n(t_1, \ldots, t_n)$, and

$$f_n(t) = \int \ldots \int h_n(\tau_1, \ldots, \tau_n)\, x(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 d\tau_n$$

In the light of these remarks, Eq. 1 can be rewritten as

$$f = \underline{H}_1[x] + \underline{H}_2[x] + \ldots + \underline{H}_n[x] + \ldots \tag{6}$$

That is, the system $\underline{H}$ has been broken into a parallel combination of systems $\underline{H}_n$, as shown in Fig. 5. This is the desired interpretation: The functional representation represents a nonlinear system as a parallel bank of systems $\underline{H}_n$ that are n^{th}-order nonlinear systems and have an impulse-response function $h_n(t_1, \ldots, t_n)$ associated with them.

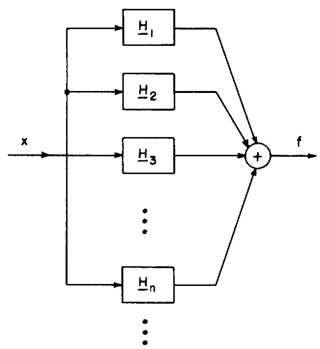


Fig. 5. Block diagram for the functional representation.

The next task is to show how these functions $h_n(t_1, \ldots, t_n)$ can be interpreted as impulse responses. The linear case is well known. If $f_1 = \underline{H}_1[x]$, and $x(t) = \delta(t+T)$, an impulse at time $-T$, then $f_1(0) = h_1(T)$, where $h_1(t)$ is the impulse response.

Now consider a generalization of the second term of the functional series

$$g_2(t) = \iint h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, y(t-\tau_2)\, d\tau_1 d\tau_2 \tag{7}$$

and represent this operationally by

$$g_2 = \underline{H}_2(xy) \tag{8}$$

This operational form will be considered in greater detail in section 2.2. The difference between $f_2 = \underline{H}_2[x]$ and $g_2 = \underline{H}_2(xy)$ should be noted. The square brackets denote an actual system operation, and the parentheses denote a mathematical operation on a pair of functions. Such a form (Eq. 8) cannot occur by itself because only single-input systems are being studied. However, it can occur in combination with other terms. Consider the system operation $f_2 = \underline{H}_2[x+y]$.

Using Eq. 4 (the actual functional relation) with Eqs. 7 and 8, we obtain

$$f_2 = \underline{H}_2(xx) + \underline{H}_2(xy) + \underline{H}_2(yx) + \underline{H}_2(yy)$$

but $h_2(t_1, t_2)$ is a symmetrical function, and so

$$f_2 = \underline{H}_2(x^2) + 2\underline{H}_2(xy) + \underline{H}_2(y^2) \tag{9}$$

where $xx = x^2$ and $yy = y^2$. In the functional form for the second-order case (Eq. 7), with $h_1(t_1, t_2) \neq h_2(t_2, t_1)$, the symmetrical function $[h_2(t_1, t_2)+h_2(t_2, t_1)]/2$ can be formed and substituted for $h_2(t_1, t_2)$ without affecting $f_2(t)$. This procedure (5) generalizes to $h_n(t_1, \ldots, t_n)$, and so it will generally be assumed in this report that $h_n(t_1, \ldots, t_n)$ is a symmetrical function in $t_1, t_2, \ldots, t_n$.

In Eq. 9, $\underline{H}_2(xy)$ has been obtained, but it is in combination with two other terms. Figure 6 shows how $\underline{H}_2(xy)$ can be isolated experimentally. If the indicated operations were performed sequentially, only one system $\underline{H}_2$ would be needed. In the system of Fig. 6, if $x(t) = \delta(t+T_1)$ and $y(t) = \delta(t+T_2)$, then the output p is $p = \underline{H}_2(xy)$, and, at time 0, $p(0) = h_2(T_1, T_2)$. This is proved by substituting these values of x(t) and y(t) in Eq. 7. Thus, $h_2(t_1, t_2)$ can be interpreted as an impulse response in a manner similar to the interpretation of the linear response $h_1(t)$. This approach can be generalized to the n^{th}-order case, and all functions $h_n(t_1, \ldots, t_n)$ may be called impulse responses. In section 4.8 we shall be concerned with measuring these impulse responses.

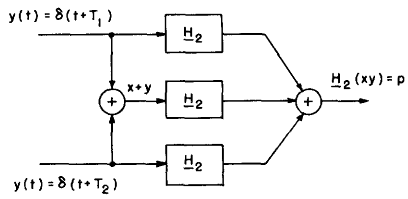


Fig. 6. Apparatus for isolating $\underline{H}_2(xy)$.

To summarize, the functional series may be regarded as representing a nonlinear system as a parallel bank of nonlinear subsystems (or operators). Each of these subsystems is specified by an impulse response, $h_n(t_1, \ldots, t_n)$.

1.6 SYSTEM TRANSFORMS

If the impulse responses $h_2(t_1, \ldots, t_n)$ are known for a system, then the output f(t), for a given input x(t), can be obtained from Eq. 1. However, the analysis of linear systems has been greatly aided by the fact that "convolution in the time domain is multiplication in the frequency domain." An analogous result holds for nonlinear systems – except that multiple-order transformations must be used.

These transforms are defined by the transform pairs:

$$Y(s_1, \ldots, s_n) = \int \cdots \int y(t_1, \ldots, t_n) \exp(s_1 t_1 + \ldots + s_n t_n)\, dt_1 \ldots dt_n \tag{10}$$

and

$$Y(t_1, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^n \int \cdots \int Y(s_1, \ldots, s_n) \exp(-s_1 t_1 - \ldots - s_n t_n)\, ds_1 \ldots ds_n \tag{11}$$

Appropriate contours of integration and values of s_1, s_2, and so on can be chosen in a manner similar to that in the linear transform case to give Fourier or Laplace transformations.

The value of the higher-order transform theory lies in the fact that

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_n)\, k_n(t_1-\tau_1, \ldots, t_n-\tau_n)\, d\tau_1 \ldots d\tau_n \tag{12}$$

has an n^{th}-order transform, $H_n(s_1, \ldots, s_n)\, K_n(s_1, \ldots, s_n)$.

Now, consider

$$f_{(2)}(t_1, t_2) = \iint h_2(\tau_1, \tau_2)\, x(t_1-\tau_1)\, x(t_2-\tau_2)\, d\tau_1 d\tau_2 \tag{13}$$

which is a special case of Eq. 12, and thus will have a transform, $F_{(2)}(s_1, s_2) = H_2(s_1, s_2)\, X(s_1)\, X(s_2)$. We are interested in the special case of Eq. 13, with $t_1 = t_2 = t$. Then

$$f_2(t) = f_{(2)}(t, t) = \iint h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \tag{14}$$

which is the second term in the functional series. Similarly, the output of an n^{th}-order system can be made artificially a function of $t_1, \ldots, t_n$, in order to take advantage of transform theory. The discussion at this point is only intended to define the transforms and indicate their possible application. In Section III we shall show how the transforms can be used to obtain the system output.

1.7 SUMMARY

We have given an introduction to the functional representation for nonlinear systems. This functional method can be used to solve a large class of nonlinear problems in which the classical methods fail, but it does have certain limitations, certainly, at the present stage of development. Furthermore, the functional representation has three very desirable properties that make it a method of considerable strength and value.

We have seen that the representation may be viewed as a parallel bank of nonlinear operations or subsystems. These subsystems are generalizations of the ordinary linear convolution operation, and are specified by impulse responses. Finally, the higher-order transform has been introduced, and its potential use indicated.

II. AN ALGEBRA OF SYSTEMS

2.1 INTRODUCTION

The second property of the functional representation is that it facilitates the combination of systems. This property was noticed by Brilliant (4), and he obtained formulas for finding the impulse responses and transforms of the component subsystems. However, these formulas are difficult to use, and do not indicate how the components of a system combine to produce the over-all system. These difficulties can be overcome by means of a representation in which the whole system can be expressed by a single equation. This representation, which is called the "Algebra of Systems," makes use of the operator system notation that was introduced in section 1.5.

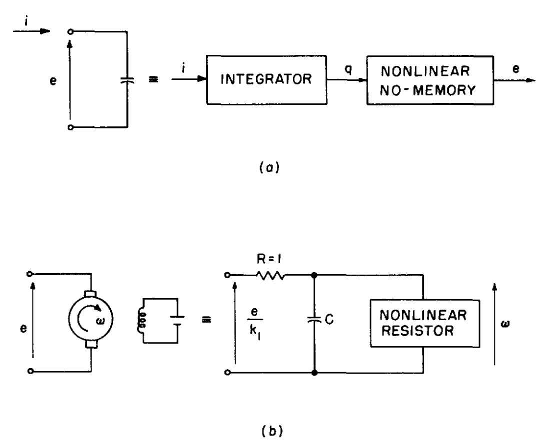


Fig. 7. Examples of nonlinear systems: (a) nonlinear capacitor; (b) dc motor.

We are primarily concerned with a certain class of physical systems. In this class, the systems are composed of:

(a) nonlinear subsystems with no memory (that is, the outputs depend on the instantaneous value of the input and are independent of the past or future values of the input);

(b) linear subsystems that, in general, have memory.

This class of systems is of a very general nature. The only class of system that appears to be definitely excluded is the hysteretic system. Two examples are shown in Fig. 7.

The nonlinear capacitor, viewed as a system, is equivalent to an integrator followed by a nonlinear no-memory operation. We can see this by considering the capacitor equation

$$e = n(q) \tag{15}$$

where e represents voltage and q, charge, and the function n represents the nonlinear relation between charge and voltage. Then

$$q(t) = \int_{-\infty}^{t} i(t)\, dt \tag{16}$$

where i(t) is the current. The block diagram of Fig. 7a shows this relation between current and voltage.

The relation between the speed ω and the armature voltage e of the dc motor is given by

$$e = k_1 \left\{ \omega + n(\omega) + k_2 \frac{d\omega}{dt} \right\} \tag{17}$$

where k_1 and k_2 are constants, and n is a function representing the nonlinear characteristic of the motor. Thus, the motor is equivalent to the circuit shown in Fig. 7, with $C = k_2$ (see Truxal (8)).

We know how to describe the linear system and the nonlinear no-memory system. The linear system can be described by its impulse response or transform, and the nonlinear no-memory system can be described by a function relating its input and output. The use of the functional representation depends on our being able to write, or approximate, this nonlinear function by a power series or a polynomial. For example, the saturating system of Fig. 2a can be approximated over a desired interval by

$$f = a_1 x + a_3 x^3 + \ldots + a_{2n+1} x^{2n+1} \tag{18}$$

The ideal clipper of Fig. 2b, on the other hand, would require an extremely large n for approximation in the form of Eq. 17. This is a practical limitation. Even very violent nonlinearities, such as the ideal clipper, can, theoretically, be very closely approximated by a polynomial.

Now the situation is: We are given a system in which the component subsystems are linear, or nonlinear no-memory, and we want to describe the over-all system by the functional representation. To do this, the subsystems (which we know how to describe) must be combined. Therefore, the ability to conveniently combine systems is very important in the use of the functional representation for system analysis.

It can be said that not only is the ability to combine nonlinear systems an important engineering problem but also that this ability is a basic need in the use of the functional representation. The algebra of systems will be developed and the relation to the system impulse responses and transforms shown.

2.2 FUNCTIONAL OPERATIONS

We introduced the operational notation in Section I. For a general system that operates on an input x(t) to produce an output f(t) (see Fig. 1), $f(t) = \underline{H}[x(t)]$ or $f = \underline{H}[x]$, where t is implicit. The system operation (Eq. 3) is denoted by $f_n = \underline{H}_n[x]$. Then, the functional series (Eq. 1) becomes $f = \underline{H}_1[x] + \underline{H}_2[x] + \ldots + \underline{H}_n[x] + \ldots$. If this form is truncated at some $\underline{H}_n[x]$, it is then a functional polynomial.

Now, if $f_n = \underline{H}_n[x]$, then

$$g_n = \underline{H}_n[\epsilon x] = \epsilon^n \underline{H}_n[x] = \epsilon^n f_n \tag{19}$$

where ϵ is a constant. If f(t) is the output of system $\underline{H}$ with input x(t), and $f_\epsilon(t)$ is the output with input $\epsilon x(t)$, it follows that

$$f_\epsilon = \epsilon \underline{H}_1[x] + \epsilon^2 \underline{H}_2[x] + \ldots + \epsilon^n \underline{H}_n[x] + \ldots \tag{20}$$

The usual Taylor, or power, series is

$$a_1\epsilon + a_2\epsilon^2 + \ldots + a_n\epsilon^n + \ldots \tag{21}$$

and comparison of Eqs. 20 and 21 shows that the functional series is very similar to a power series. It will be shown in section 6.6 that there is a strong mathematical connection between them. This relationship serves to relate the functional series to ordinary mathematical series.

We have represented the generalized second-order operation

$$g_2(t) = \iint h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, y(t-\tau_2)\, d\tau_1 d\tau_2 \tag{22}$$

by

$$g_2 = \underline{H}_2(xy) \tag{23}$$

When x = y we have, $g_2 = \underline{H}_2(xx) = \underline{H}_2(x^2)$, and since this represents a real input into the system $\underline{H}_2$,

$$g_2 = \underline{H}_2(x^2) = \underline{H}_2[x] \tag{24}$$

Terms of the form of Eq. 24 do not occur alone, but in combination with other terms. If $f = \underline{H}_2[x+y]$, then from the definition of $\underline{H}_2$,

$$\begin{aligned} f(t) &= \iint h_2(\tau_1, \tau_2)\{x(t-\tau_1)+y(t-\tau_2)\}\{x(t-\tau_2)+y(t-\tau_2)\}\, d\tau_1 d\tau_2 \\ &= \iint h(\tau_1, \tau_2)\{x(t-\tau_1)x(t-\tau_2)+x(t-\tau_1)y(t-\tau_2)+y(t-\tau_1)x(t-\tau_2)+y(t-\tau_1)y(t-\tau_2)\}\, d\tau_1 d\tau_2 \end{aligned} \tag{25}$$

But, since $h(\tau_1, \tau_2)$ is symmetrical,

$$\begin{aligned} f(t) &= \iint h(\tau_1, \tau_2)\{x(t-\tau_1)x(t-\tau_2)+2x(t-\tau_1)y(t-\tau_2)+y(t-\tau_1)y(t-\tau_2)\}\, d\tau_1 d\tau_2 \\ &= \iint h(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 + 2\iint h(\tau_1, \tau_2)\, x(t-\tau_1)\, y(t-\tau_2)\, d\tau_1 d\tau_2 \\ &\quad + \iint h(\tau_1, \tau_2)\, y(t-\tau_1)\, y(t-\tau_2)\, d\tau_1 d\tau_2 \end{aligned} \tag{26}$$

In terms of the definitions of Eqs. 22 and 23, Eq. 26 can be written

$$f = \underline{H}_2(x^2) + 2\underline{H}_2(xy) + \underline{H}_2(y^2) \tag{27}$$

This expansion of $\underline{H}_2[x+y]$ can be obtained directly in the short notation, by the following sequence of steps:

$$f = \underline{H}_2[x+y]$$

$$= \underline{H}_2((x+y)^2)$$

$$= \underline{H}_2(x^2+2xy+y^2)$$

$$= \underline{H}_2(x^2) + 2\underline{H}_2(xy) + \underline{H}_2(y^2)$$

$$= \underline{H}_2[x] + 2\underline{H}_2(xy) + \underline{H}_2[y]$$

and this is validated by Eq. 27. Thereby, the form $\underline{H}_2(xy)$ occurs in combination with other similar forms.

Similarly, for the third-order case, $f_3 = \underline{H}_3[x+y] = \underline{H}_3((x+y)^3) = \underline{H}_3(x^3+3x^2y+3xy^2+y^2)$, or $f_3 = \underline{H}_3(x^3) + 3\underline{H}_3(x^2y) + 3\underline{H}_3(xy^2) + \underline{H}_3(y^3)$. This directly generalizes for the n^{th}-order case. Not only is this a useful interpretation of the functional operation, but it will also be shown, in the course of this report, to be extremely useful for dealing with inputs that are composed of sums of simple functions such as sinusoids. Also, this is of great importance in the algebraic expansion used for determining the system impulse responses and transforms.

We have now accomplished two aims:

(a) The notion of functional power series has been introduced.

(b) The concept of nonlinear operations has been defined as generalized multiplication operations on multiple signals. For example, $\underline{H}_3(xyz)$ is an operation on a triplet of functions x(t), y(t), and z(t).

2.3 SYSTEM COMBINATIONS

There are three basic means of combining nonlinear systems — addition, multiplication, and cascading. The addition combination of two systems involves putting the same input into the two systems and combining the two outputs in an adder. This is shown in Fig. 8a and is written algebraically: $\underline{L} = \underline{J} + \underline{K}$, where $\underline{L}$ is the combined system, and $\underline{J}$ and $\underline{K}$ are the component systems.

The multiplication combination is similar, except that a multiplier is substituted for the adder. The diagram is shown in Fig. 8b and the combination is written

$$\underline{L} = \underline{J} \cdot \underline{K} \tag{28}$$

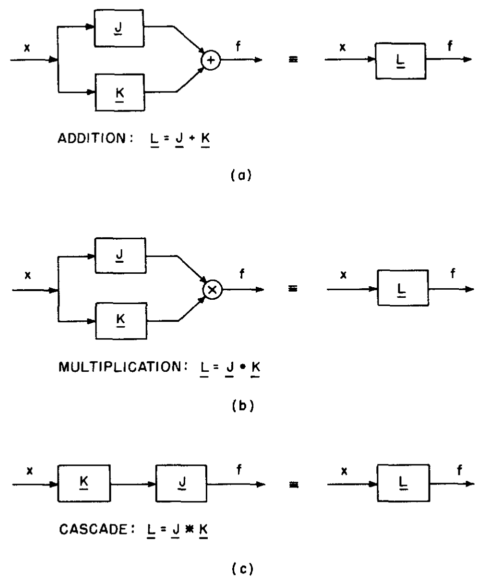


Fig. 8. System combinations.

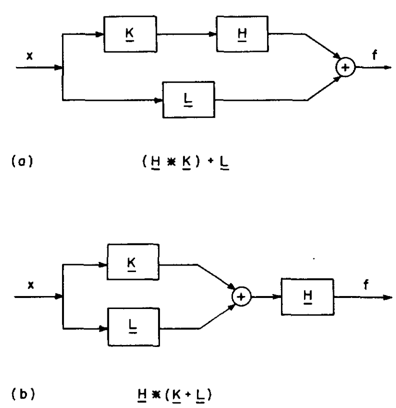


Fig. 9. Illustrating the use of brackets.

In the cascade combination the output of one system is the input of the other. This is shown in Fig. 8c and is written L = J * K. Expressing it in words, we can use "plus" for +, "times" for ·, and "cascade" for *. Then, for example, J + K is read, "jay plus kay."

It is convenient to have a bracketing operation, in addition to the other operations. This is used to remove ambiguity from the algebraic expressions. For example, the system (J*K) + L is the cascade system J * K plus the system L. This is shown in Fig. 9a. However, the system J * (K+L) is the system J cascaded with the system (K+L). This combination is shown in Fig. 9b. The bracket, then, has the same grouping meaning that it usually has in algebra, and all terms in parentheses specify a composite system.

For the system operation $f = \underline{L}[x]$, where $\underline{L} = \underline{J} + \underline{K}$, we can write

$$f = (\underline{J}+\underline{K})[x] \tag{29}$$

Similarly, if $\underline{L} = \underline{J} * \underline{K}$, we can write

$$f = (\underline{J}*\underline{K})[x] \tag{30}$$

Equation 30, however, has another form. Let

$$y = \underline{K}[x] \tag{31}$$

Then, by the definition of the cascade operation (see Fig. 8c),

$$f = \underline{J}[y] \tag{32}$$

Substitution of Eq. 32 in Eq. 31 yields

$$f = \underline{J}[K[x]] \tag{33}$$

as an alternative form for Eq. 30.

Now that we have the basic definitions of this algebra, we shall proceed to develop it. In view of the addition definition, the functional representation is seen to be an expansion of a system $\underline{H}$, and so $\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n + \dots$.

Now, this algebra will have two uses:

(a) To expand a system in terms of its component linear and nonlinear no-memory subsystems.

(b) To allow block-diagram manipulation.

In order to perform these manipulations, or rearrangements, certain algebraic rules must be developed. For the addition and multiplication operations the rules are similar to those usually followed in algebra. The rules for the cascade operation are somewhat different. These rules will be given in the form of eight axioms. The proofs are based on the physical significance of the algebraic operations.

The first two axioms are concerned with the addition operation.

Axiom 1. $\underline{J} + \underline{K} = \underline{K} + \underline{J}$ (34)

This combination is illustrated in Fig. 8a. Axiom 1 states that both $\underline{J} + \underline{K}$ and $\underline{K} + \underline{J}$ stand for the additive combination of Fig. 8a.

Axiom 2. $\underline{J} + (\underline{K}+\underline{L}) = (\underline{J}+\underline{K}) + \underline{L}$ (35)

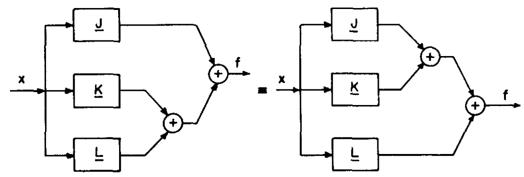
This axiom is illustrated by Fig. 10a. The diagram shows that it does not matter whether $\underline{K}$ and $\underline{L}$ or $\underline{J}$ and $\underline{K}$ are grouped together.

The next two axioms are like axioms 1 and 2, except that they have plus replaced by times.

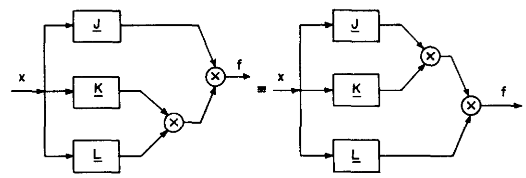
Axiom 3. $\underline{J} \cdot \underline{K} = \underline{K} \cdot \underline{J}$ (36)

Axiom 4. $\underline{J} \cdot (\underline{K}\cdot\underline{L}) = (\underline{J}\cdot\underline{K}) \cdot \underline{L}$ (37)

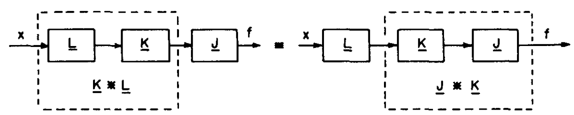
The diagram for the axiom 3 combination is Fig. 8b. Axiom 3 states that both $\underline{J} \cdot \underline{K}$



(a) $\underline{J} + (\underline{K} + \underline{L}) = (\underline{J} + \underline{K}) + \underline{L}$

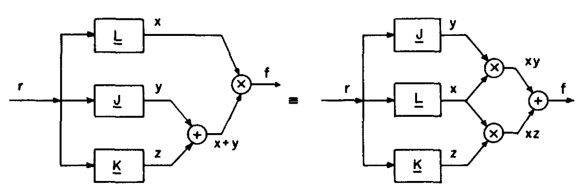


(b) $\underline{J} \bullet (\underline{K} \bullet \underline{L}) = (\underline{J} \bullet \underline{K}) \bullet \underline{L}$

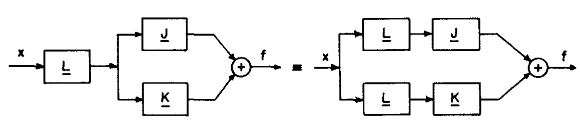


(c) $\underline{J} * (\underline{K} * \underline{L}) = (\underline{J} * \underline{K}) * \underline{L}$

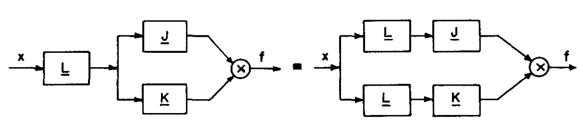
Fig. 10. Illustration of axioms.



(a) $\underline{L} \bullet (\underline{H} + \underline{K}) = (\underline{L} \bullet \underline{H}) + (\underline{L} \bullet \underline{K})$



(b) $(\underline{J} + \underline{K}) * \underline{L} = (\underline{J} * \underline{L}) + (\underline{K} * \underline{L})$



(c) $(J \bullet K) * L = (J * L) \bullet (K * L)$

Fig. 11. Illustration of axioms.

and $\underline{K} \cdot \underline{J}$ stand for this combination. Figure 10b is the diagram for axiom 4. It does not matter whether $\underline{K}$ and $\underline{L}$ or $\underline{J}$ and $\underline{K}$ are grouped together.

The last axiom of this group concerns the cascade operation

Axiom 5. $\underline{J} * (\underline{K}*\underline{L}) = (\underline{J}*\underline{K}) * \underline{L}$ (38)

This axiom is illustrated by Fig. 10c, where it is shown that the ( ) operation has no physical significance. It is simply a matter of algebraic convenience.

Then, there are three axioms dealing with combined operations.

Axiom 6. $\underline{L} \cdot (\underline{J}+\underline{K}) = (\underline{L}\cdot\underline{J}) + (\underline{L}\cdot\underline{K})$ (39)

The diagram for this axiom is Fig. 11a. Axiom 6 is true because

$$f = x(y+z) = xy + yz \qquad (40)$$

where x, y, and z are the outputs of $\underline{L}$, $\underline{J}$, and $\underline{K}$, respectively.

A similar axiom holds for the plus and cascade combination.

Axiom 7. $(\underline{J}+\underline{K}) * \underline{L} = (\underline{J}*\underline{L}) + (\underline{K}*\underline{L})$ (41)

This is shown in Fig. 11b; the two systems illustrated there are equivalent

Axiom 8. $(\underline{J}\cdot\underline{K}) * \underline{L} = (\underline{J}*\underline{L}) \cdot (\underline{K}*\underline{L})$ (42)

The two equivalent systems for this axiom are shown in Fig. 11c.

It is also important to know which rearrangements are not legitimate. In particular, we note that, in general,

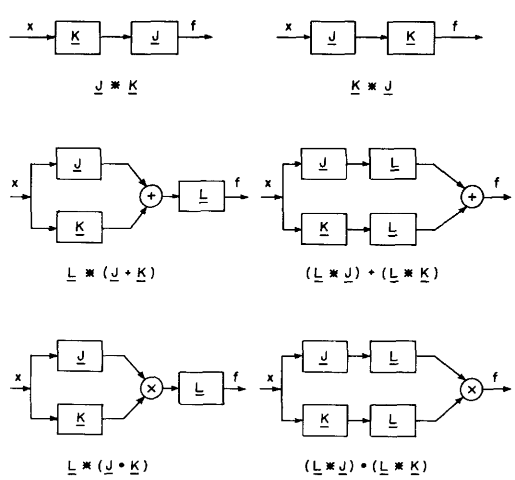


Fig. 12. Illustration of combinations.

$$\underline{J} * \underline{K} \neq \underline{K} * \underline{J}$$

$$\underline{L} * (\underline{J}+\underline{K}) \neq (\underline{L}*\underline{J}) + (\underline{L}*\underline{K})$$

$$\underline{L} * (\underline{J}\cdot\underline{K}) \neq (\underline{L}*\underline{J}) \cdot (\underline{L}*\underline{K}) \tag{43}$$

Block diagrams for various expressions are given in Fig. 12, and these relations will be demonstrated by means of simple counterexamples. Let $\underline{J}[x] = ax^2$, $\underline{K}[x] = bx^2$, and $\underline{L}[x] = cx^2$, where a, b, and c are constants. Then $(\underline{J}*\underline{K})[x] = a(bx^2)^2 = ab^2x^4$, and $(\underline{K}*\underline{J})[x] = b(ax^2)^2 = a^2bx^4$, with the result that $(\underline{J}*\underline{K})[x] \neq (\underline{K}*\underline{J})[x]$, and thus Eq. 43 is established in this special case. We also have

$$(\underline{L}*(\underline{J}+\underline{K}))[x] = c(ax^2+bx^2)^2 = c(a+b)^2\ x^4 \tag{44}$$

and

$$((\underline{L}*\underline{J})+(\underline{L}*\underline{K}))[x] = c(ax^2)^2 + c(bx^2)^2 = c(a^2+b^2)\ x^4 \tag{45}$$

Since Eqs. 44 and 45 are not equal, Eq. 43 has been justified. Now

$$(\underline{L}*(\underline{J}\cdot\underline{K}))[x] = c(ax^2\cdot bx^2)^2 = ca^2b^2x^8$$

and

$$((\underline{L}*\underline{J})\cdot(\underline{L}*\underline{K}))[x] = c(ax^2)^2 \cdot c(bx^2)^2 = c^2a^2b^2x^8$$

and so Eq. 43 is valid.

There are, however, two important special cases:

$$\underline{J}_1 * \underline{K}_1 = \underline{K}_1 * \underline{J}_1 \tag{46}$$

and

$$\underline{L}_1 * (\underline{J}+\underline{K}) = (\underline{L}_1*\underline{J}) + (\underline{L}_1*\underline{K}) \tag{47}$$

Equation 46 is known from the theory of linear systems (9). To prove Eq. 47, let

$$\underline{J}[x] = y \tag{48}$$

and

$$\underline{K}[x] = z \tag{49}$$

then $(\underline{L}_1*(\underline{J}+\underline{K}))[x] = \underline{L}_1[y+z]$. But $\underline{L}_1$ is a linear system, and by superposition, $\underline{L}_1[y+z] = \underline{L}_1[y] + \underline{L}_1[z]$. Substituting Eqs. 48 and 49 in this expression gives

$$\underline{L}_1[\underline{J}[x]+\underline{K}[x]] = \underline{L}_1[\underline{J}[x]] + \underline{L}_1[\underline{K}[x]]$$

or

$$\underline{L}_1 * (\underline{J}+\underline{K}) = (\underline{L}_1*\underline{J}) + (\underline{L}_1*\underline{K})$$

and Eq. 47 is proved.

2.4 ORDER OF SYSTEMS

As we have mentioned, the functional representation expands a system $\underline{H}$ in a series

$$\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n + \dots \tag{50}$$

In section 1.5 $\underline{H}_1$ was defined as a linear system, $\underline{H}_2$ as a quadratic system, and so on. $\underline{H}_n$ is called an n^{th}-order system, and $\underline{H}_n[\epsilon x] = \epsilon^n \underline{H}_n[x]$, where x is the input signal, and ϵ is a constant. Equation 50 shows that this order differentiates between the terms in the functional representation; that is, the first term is linear (first-order), the second is quadratic (second-order), and so on. It is possible for a system to have a dc bias at the output which is unaffected by the input. This bias can be taken as the result of a zero-order system $\underline{H}_o$ with the property that

$$\underline{H} = \underline{H}_o + \underline{H}_1 + \dots + \underline{H}_n + \dots$$

where $\underline{H}_o$ is specified by a constant h_o. However, since $\underline{H}_o$ does not have any input-output relation, we shall usually not include it in the functional series.

So that a combined system can be expanded in the functional series (Eq. 50), the effect of combinations on order must be noted. The system $\underline{L}$, with

$$\underline{L} = \underline{A}_n + \underline{B}_m \tag{51}$$

contains both n^{th}- and m^{th}-order parts, as Eq. 51 shows. The system $\underline{K}$, with

$$\underline{K} = \underline{A}_n \cdot \underline{B}_m \tag{52}$$

is a system of order m + n. This order follows because

$$\underline{K}[\epsilon x] = \underline{A}_n[\epsilon x] \cdot \underline{B}_m[\epsilon x] = \epsilon^{m+n} \underline{K}[x]$$

The cascade system $\underline{H}$, with

$$\underline{H} = \underline{A}_n * \underline{B}_m \tag{53}$$

is a system of order mn. This is shown by

$$\underline{H}[\epsilon x] = \underline{A}_n[\underline{B}_m[\epsilon x]] = (\epsilon^m)^n \, \underline{A}_n[\underline{B}_m[x]] = \epsilon^{mn} \, \underline{H}[x]$$

in which we have used the alternative cascade definition (Eq. 33).

Now that the effect of system combination upon ordering has been explained, it is possible to expand a combined system in the functional series or the functional polynomial. Before giving an example of a combined system, several special systems will be considered.

2.5 SPECIAL SYSTEMS

We shall now introduce the notation for some special systems. The nonlinear no-memory system will usually be denoted by $\underline{N}$, so that $f = \underline{N}[x]$. In polynomial or power-series form,

$$\underline{N} = \underline{N}_1 + \underline{N}_2 + \dots + \underline{N}_m + \dots \tag{54}$$

and then

$$f = n_1 x + n_2 x^2 + \dots + n_m x^m + \dots \tag{55}$$

A particular linear no-memory system is the identity system $\underline{I}$, which has the definition $x = \underline{I}[x]$.

The zero system $\underline{0}$ is defined as

$$0 = \underline{0}[x] \tag{56}$$

In algebraic equations, 0 will be used to denote the system $\underline{0}$.

These rather obvious properties should be noted:

$$\underline{H} + \underline{0} = \underline{H}$$

and

$$\underline{I} * \underline{H} = \underline{H} * \underline{I} = \underline{H}$$

In this algebra it is often convenient to replace the nonlinear no-memory operations by multiplication operations. To do this, consider the term $\underline{N}_m * \underline{H}$. By virtue of the definition of $\underline{N}$ given by Eqs. 54 and 55,

$$\underline{N}_m[x] = n_m x^m \tag{57}$$

Now, if $x = \underline{H}[y]$, then $\underline{N}_m[x] = \underline{N}_m[\underline{H}[y]] = (\underline{N}_m * \underline{H})[y]$, and from Eq. 57, $\underline{N}_m[x] = n_m(\underline{H}[y])^m$. By definition of the multiplication operation (Eq. 28) of this algebra, this procedure gives

$$\underline{N}_m * \underline{H} = n_m \underbrace{\underline{H} \cdot \underline{H} \cdot \dots \cdot \underline{H}}_{m \text{ times}} \tag{58}$$

Then, if we define

$$\underbrace{\underline{H} \cdot \underline{H} \cdot \dots \cdot \underline{H}}_{m \text{ times}} = \underline{H}^m$$

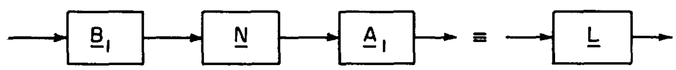
we have

$$\underline{N}_m * \underline{H} = n_m \underline{H}^m \tag{59}$$

where n_m is just a gain constant. The no-memory system $\underline{N}_m$ has been replaced by a multiplication operation thereby, and a sum of no-memory systems $\underline{H}_m$ can be replaced by a sum of multiplication operations.

2.6 EXAMPLE 1.

The combined system will now be illustrated by an example. Let us consider the system of Fig. 13 in which $\underline{L} = \underline{A}_1 * \underline{N} * \underline{B}_1$. This system can be viewed as an ampli-



$\underline{L} = \underline{A}_1 * \underline{N} * \underline{B}_1$

Fig. 13. Illustrative cascade system.

fier with nonlinear distortion. $\underline{A}_1$ and $\underline{B}_1$ are linear systems and $\underline{N}$ is a nonlinear no-memory system. Let $\underline{N}$ have a linear and a cubic part, so that $\underline{N} = \underline{N}_1 + \underline{N}_3$. Then

$$\underline{L} = \underline{A}_1 * (\underline{N}_1 + \underline{N}_3) * \underline{B}_1$$

and by using axiom 7 (Eq. 41), we obtain

$$\underline{L} = \underline{A}_1 * (\underline{N}_1 * \underline{B}_1 + \underline{N}_3 * \underline{B}_1)$$

By use of Eq. 47, we have

$$\underline{L} = \underline{A}_1 * \underline{N}_1 * \underline{B}_1 + \underline{A}_1 * \underline{N}_3 * \underline{B}_1$$

When $\underline{N}_1$ and $\underline{N}_3$ are replaced by multiplication operations, we have

$$\underline{L} = \underline{A}_1 * (n_1 \underline{B}_1) + \underline{A}_1 * \left(n_3 \underline{B}_1^3\right)$$

or

$$\underline{L} = n_1 \underline{A}_1 * \underline{B}_1 + n_3 \underline{A}_1 * \underline{B}_1^3$$

since $\underline{A}_1$ is linear. Now $\underline{L} = \underline{L}_1 + \underline{L}_3$, where $\underline{L}_1 = n_1 \underline{A}_1 * \underline{B}_1$, and $\underline{L}_3 = n_3 \underline{A}_1 * \underline{B}_1^3$.

This example illustrates how this algebra can be used to describe a system in terms of its component subsystems. Next, we want to relate the algebraic representation to the system impulse responses or transforms. Once this is done, we can proceed to find the system response to various excitations. But, first, two other topics in this algebra must be considered.

2.7 CASCADE OPERATIONS

Strictly speaking, the cascade operations involved in combining these linear subsystems and no-memory nonlinear subsystems will not involve cascading nonlinear systems with memory. However, algebraic simplification is often obtained by grouping a number of subsystems to produce a composite subsystem that is nonlinear and has memory (see sec. 2.8 for an illustration of this point). This section is concerned with nonlinear systems with memory, in cascade.

The cascade system $\underline{A}_n * \underline{B}_m$ has been shown to be of order mn. Now consider the system $\underline{L}$, in which

$$\underline{L} = \underline{A}_2 * (\underline{B}_n + \underline{C}_m) \tag{60}$$

To determine the order of this system, we shall develop an expansion for $\underline{A}_2 * (\underline{B}_n + \underline{C}_m)$. Now

$$\underline{L}[x] = (\underline{A}_2 * (\underline{B}_n + \underline{C}_m))[x] = \underline{A}_2[\underline{B}_n[x] + \underline{C}_m[x]]$$

Let

$$y = \underline{B}_n[x] \tag{61a}$$

$$z = \underline{C}_m[x] \tag{61b}$$

and then

$$\underline{L}[x] = \underline{A}_2[y+z]$$

$$= \underline{A}_2((y+z)^2)$$

$$= \underline{A}_2(y^2) + 2\underline{A}_2(yz) + \underline{A}_2(z^2)$$

Now, substitution of Eqs. 61a and 61b gives

$$\underline{L}[x] = \underline{A}_2\left((\underline{B}_n[x])^2\right) + 2\underline{A}_2(\underline{B}_n[x] \cdot \underline{C}_m[x]) + \underline{A}_2\left((\underline{C}_m[x])^2\right)$$

Then if we define

$$(\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m))[x] = \underline{A}_2(\underline{B}_m[x] \cdot \underline{C}_m[x])$$

with the use of the operation "o", the system $\underline{L}$ becomes

$$\underline{L} = \underline{A}_2 \circ \left(\underline{B}_n^2\right) + 2\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m) + \underline{A}_2 \circ \left(\underline{C}_m^2\right) \tag{62}$$

(Note that $\underline{A}_2 \circ \left(\underline{B}_n^2\right) = \underline{A}_2 * \underline{B}_n$.)

Now that Eq. 62 has been established, we see that it can be quickly obtained from Eq. 60, as follows:

$$\underline{L} = \underline{A}_2 * (\underline{B}_n + \underline{C}_m)$$

$$= \underline{A}_2 \circ (\underline{B}_n + \underline{C}_m)^2$$

$$= \underline{A}_2 \circ \left(\underline{B}_n^2 + 2\underline{B}_n \cdot \underline{C}_m + \underline{C}_m^2\right)$$

$$= \underline{A}_2 \circ \left(\underline{B}_n^2\right) + 2\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m) + \underline{A}_2 \circ \left(\underline{C}_m^2\right) \tag{63}$$

The term $\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m)$ is an operator of order m + n because

$$(\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m))[\epsilon x] = \underline{A}_2(\underline{B}_n[\epsilon x] \cdot \underline{C}_m[\epsilon x])$$

$$= \underline{A}_2\left(\epsilon^{m+n} \underline{B}_n[x] \cdot \underline{C}_m[x]\right)$$

$$= \epsilon^{m+n} \underline{A}_2(\underline{B}_n[x] \cdot \underline{C}_m[x])$$

$$= \epsilon^{m+n}(\underline{A}_2 \circ (\underline{B}_n \cdot \underline{C}_m))[x] \tag{64}$$

Therefore, Eq. 63 shows that $\underline{A}_2 * (\underline{B}_n + \underline{C}_m)$ can be expanded into three operators of order 2n, m + n, and 2m.

The case $\underline{A}_3 * (\underline{B}_n + \underline{C}_m)$ can be expanded in a similar manner:

$$\underline{A}_3 * (\underline{B}_n + \underline{C}_m) = \underline{A}_3 \circ (\underline{B}_n + \underline{C}_m)^3 = \underline{A}_3 \circ \left(\underline{B}_n^3\right) + 3\underline{A}_3 \circ \left(\underline{B}_n^2 \cdot \underline{C}_m\right)$$

$$+ 3\underline{A}_3 \circ \left(\underline{B}_n \cdot \underline{C}_m^2\right) + \underline{A}_3 \circ \left(\underline{C}_m^3\right)$$

where $\underline{A}_3 \circ \left(\underline{B}_n^2 \cdot \underline{C}_m\right)$ is of order 2n + m, and

$$\left(\underline{A}_3 \circ \left(\underline{B}_n^2 \cdot \underline{C}_m\right)\right)[x] = \underline{A}_3\left(\underline{B}_n^2[x] \cdot \underline{C}_m[x]\right)$$

This expansion of the cascade operation can be generalized to any order. For example,

$$\underline{A}_s * (\underline{B}_n + \underline{C}_m + \ldots + \underline{P}_r) = \underline{A}_s \circ (\underline{B}_n + \underline{C}_m + \ldots + \underline{P}_r)^s$$

and has a typical term in its expansion:

$$\underline{A}_s \circ \underbrace{(\underline{B}_n \cdot \underline{C}_m \cdot \ldots)}_{s \text{ terms}}$$

which is of order n + m + ...

In this manner, a cascade combination of systems can be split up into a sum of single operations. Each of these simple operations has a single transform of impulse response associated with it, which will be given later.

2.8 FEEDBACK SYSTEMS

Example 1 was for a feed-through system. Therefore, obtaining its functional expansion was a straightforward procedure. We shall now develop the procedure for determining the functional expansion for a feedback system. The single-loop feedback system is shown in Fig. 14a, in which $\underline{A}$ and $\underline{B}$ are nonlinear systems that have a known functional expansion. Figure 14b is an equivalent system, in which the feedback system of Fig. 14a has been split into the system $\underline{A}$ cascaded with a simpler feedback system.

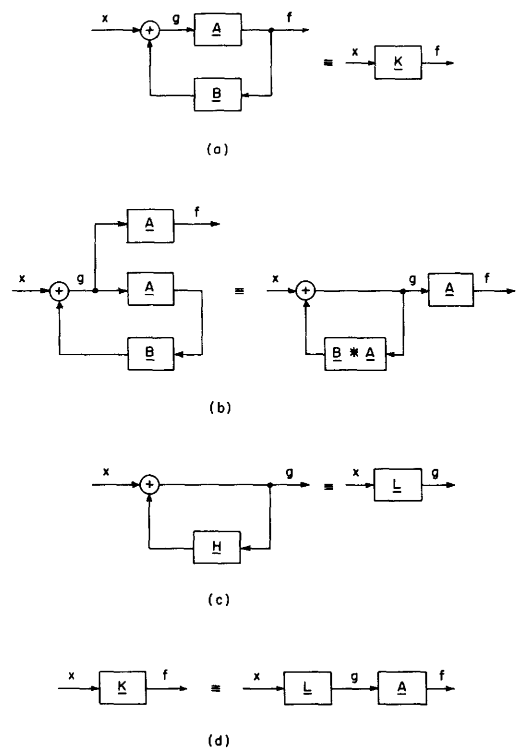


Fig. 14. (a) Nonlinear feedback system. (b) Equivalent system. (c) System $\underline{L}$. (d) Combination of $\underline{A}$ and $\underline{L}$.

Let $\underline{B} * \underline{A} = \underline{H}$, and let the simpler feedback system be denoted explicitly by $\underline{L}$, as shown in Fig. 14c. Then the feedback system of Fig. 14a, which is explicitly denoted by $\underline{K}$, is given by

$$\underline{K} = \underline{A} * \underline{L} \tag{65}$$

as shown in Fig. 14d. Since $\underline{A}$ is known, $\underline{K}$ can be obtained from Eq. 65, once $\underline{L}$ has been determined. We shall determine $\underline{L}$ first and then find $\underline{K}$ from Eq. 65, because this is easier than developing $\underline{K}$ directly. (In many problems $\underline{K}$ can be found directly. In this general case, such a procedure is difficult.)

For the feedback system $\underline{L}$, output g is related to input x by

$$g = x + \underline{H}[g] \tag{66}$$

which relates g implicitly to x. However, it is desired to have an explicit relation

$$g = \underline{L}[x] \tag{67}$$

and so, if we substitute Eq. 67 in Eq. 66, we have

$$\underline{L}[x] = \underline{x} + \underline{H}[\underline{L}[x]]$$

Writing this as a system equation, we obtain

$$\underline{L} = \underline{I} + \underline{H} * \underline{L} \tag{68}$$

where $\underline{I}$ is the identity system. Equation 68 is an implicit equation for $\underline{L}$. Now, we have assumed that $\underline{A} = \underline{A}_1 + \underline{A}_2 + \dots + \underline{A}_n + \dots$ and $\underline{B} = \underline{B}_1 + \underline{B}_2 + \dots + \underline{B}_n + \dots$. Therefore the expansion

$$\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n + \dots \tag{69}$$

is known, since $\underline{H} = \underline{B} * \underline{A}$.

Now, we desire to find $\underline{L}$ in the series

$$\underline{L} = \underline{L}_1 + \underline{L}_2 + \dots + \underline{L}_n + \dots \tag{70}$$

Therefore, Eqs. 69 and 70 are substituted in the system equation (Eq. 68), and

$$\underline{L}_1 + \underline{L}_2 + \underline{L}_3 + \dots = \underline{I} + (\underline{H}_1+\underline{H}_2+\underline{H}_3+\dots) * (\underline{L}_1+\underline{L}_2+\underline{L}_3+\dots) \tag{71}$$

Now the $\underline{L}_n$ can be found in terms of the $\underline{H}_n$ by equating the n^{th}-order system on the left-hand side of Eq. 71 to the n^{th}-order system on the right-hand side. So that the order can be recognized, Eq. 71 must be expanded as follows:

$$\begin{aligned}\underline{L}_1 + \underline{L}_2 + \underline{L}_3 + \dots = \underline{I} &+ (\underline{H}_1*\underline{L}_1+\underline{H}_1*\underline{L}_2+\underline{H}_1*\underline{L}_3+\dots) \\ &+ \left(\underline{H}_2\circ\left(\underline{L}_1^2\right)+2\underline{H}_2\circ(\underline{L}_1\cdot\underline{L}_2)+\underline{H}_2\circ\left(\underline{L}_2^2\right)+\dots\right) \\ &+ \left(\underline{H}_3\circ\left(\underline{L}_1^3\right)+3\underline{H}_3\circ\left(\underline{L}_1^2\cdot\underline{L}_2\right)+3\underline{H}_3\circ\left(\underline{L}_1\cdot\underline{L}_2^2\right) \right. \\ &\left. + H_3\circ\left(L_2^3\right)+\dots\right)+\dots\end{aligned}$$

Equating equal orders then yields:

$$\underline{L}_1 = \underline{I} + \underline{H}_1 * \underline{L}_1 \tag{72}$$

$$\underline{L}_2 = \underline{H}_1 * \underline{L}_2 + \underline{H}_2 \circ \left(\underline{L}_1^2\right) \tag{73}$$

$$\underline{L}_3 = \underline{H}_1 * \underline{L}_3 + 2\underline{H}_2 \circ (\underline{L}_1\cdot\underline{L}_2) + \underline{H}_3 \circ \left(\underline{L}_1^3\right) \tag{74}$$

and so on.

By way of explanation, if $y = \underline{A}[x]$ and $z = -y$, then $z = -\underline{A}[x]$. Now, if $g = \underline{B}[x]$, then $f = g + z = \underline{B}[x] + (-\underline{A}[x])$. Taking $f = \underline{H}[x] = \underline{B}[x] + (-\underline{A}[x])$ gives a system equation

$$\underline{H} = \underline{B} + (-\underline{A}) \text{ or } \underline{H} = \underline{B} - \underline{A}$$

This defines the minus sign in this algebra. The minus, or subtraction, operation obeys all the rules for the addition operation. Thus by subtracting $\underline{H}_1 * \underline{L}_1$ from both sides of Eq. 72, we have

$$\underline{L}_1 - (\underline{H}_1 * \underline{L}_1) = \underline{I} + (\underline{H}_1 * \underline{L}_1) - (\underline{H}_1 * \underline{L}_1)$$

or

$$\underline{L}_1 - (\underline{H}_1 * \underline{L}_1) = \underline{I}$$

or

$$(\underline{I} - \underline{H}_1) * \underline{L}_1 = \underline{I} \tag{75}$$

because $\underline{I} * \underline{L}_1 = \underline{L}_1$. Equation 75 is, then, an alternative form of Eq. 72. In a similar manner, Eq. 73 becomes

$$(\underline{I} - \underline{H}_1) * \underline{L}_2 = \underline{H}_2 \circ \left(\underline{L}_1^2\right) \tag{76}$$

and Eq. 74 becomes

$$(\underline{I} - \underline{H}_1) * \underline{L}_3 = 2\underline{H}_2 \circ (\underline{L}_1 \cdot \underline{L}_2) + \underline{H}_3 \circ \left(\underline{L}_1^3\right) \tag{77}$$

Now, if we precascade Eq. 75 (formal justification will be given in Sec. VI) by the inverse of $(\underline{I} - \underline{H}_1)$, which is denoted $(\underline{I} - \underline{H}_1)^{-1}$, then

$$(\underline{I} - \underline{H}_1)^{-1} * (\underline{I} - \underline{H}_1) * \underline{L}_1 = (\underline{I} - \underline{H}_1)^{-1} \tag{78}$$

But $(\underline{I} - \underline{H}_1)^{-1}$ is the inverse of $(\underline{I} - \underline{H}_1)$, and so $(\underline{I} - \underline{H}_1)^{-1} * (\underline{I} - \underline{H}_1) = \underline{I}$, and Eq. 78 becomes

$$\underline{L}_1 = (\underline{I} - \underline{H}_1)^{-1} \tag{79}$$

(If $y = \underline{H}[x]$, then there is a $\underline{K}$ for which $x = \underline{K}[y]$. This $\underline{K}$ is the inverse of $\underline{H}$ and we shall denote $\underline{K}$ by $\underline{H}^{-1}$. The inverse is considered in more detail in sec. 6.3. The inverse of a linear system is well defined in linear theory.)

Similarly, Eq. 76 becomes

$$\underline{L}_2 = (\underline{I} - \underline{H}_1)^{-1} * \left(\underline{H}_2 \circ \left(\underline{L}_1^2\right)\right)$$

and Eq. 77 becomes

$$\underline{L}_3 = (\underline{I} - \underline{H}_1)^{-1} * \left(2\underline{H}_2 \circ (\underline{L}_1 \cdot \underline{L}_2) + \underline{H}_3 \circ \left(\underline{L}_1^3\right)\right)$$

In this manner, the $\underline{L}_n$ can be found for the feedback system $\underline{L}$.

The functional series for the feedback system $\underline{K}$ is then given by

$$\underline{K} = \underline{A} * \underline{L}$$

$$= (\underline{A}_1+\underline{A}_2+\ldots) * (\underline{L}_1+\underline{L}_2+\ldots)$$

$$= \underline{A}_1 * \underline{L}_1 + \underline{A}_2 \circ \left(\underline{L}_1^2\right) + 2\underline{A}_2 \circ (\underline{L}_1 \cdot \underline{L}_2) + \underline{A}_3 \circ \left(\underline{L}_1^3\right) + \ldots$$

and thus

$$\underline{K}_1 = \underline{A}_1 * \underline{L}_1 \tag{80}$$

$$\underline{K}_2 = \underline{A}_2 \circ \left(\underline{L}_1^2\right) \tag{81}$$

$$\underline{K}_3 = 2\underline{A}_2 \circ (\underline{L}_1 \cdot \underline{L}_2) + \underline{A}_3 \circ \left(\underline{L}_1^3\right) \tag{82}$$

and so on. The validity of the series expansion

$$\underline{K} = \underline{K}_1 + \underline{K}_2 + \ldots + \underline{K}_n + \ldots \tag{83}$$

will be considered in Section VI, but it may be said now that it is generally rapidly convergent for sufficiently bounded input.

In any particular problem there are two alternatives. We could use the equations for $\underline{K}_n$ for the general case of Fig. 14a (the first three equations are Eqs. 80, 81, and 82), and substitute the particular $\underline{A}$ and $\underline{B}$ that are being used. A better procedure is

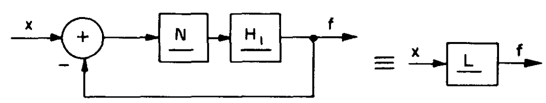


Fig. 15. Nonlinear servo system.

to work out the $\underline{K}_n$, by the method just described, for each particular case. This is not too difficult after some practice.

As an example of this method, consider the feedback system of Fig. 15. In this case

$$\underline{L} = \underline{H}_1 * \underline{N} * (\underline{I}-\underline{L}) \tag{84}$$

where $\underline{H}_1$ is a linear system, and $\underline{N} = \underline{I} + \underline{N}_3$.

This system is sufficiently simple that the series for $\underline{L}$ can be obtained directly. Equation 84 can be rewritten as

$$\underline{L} = \underline{H}_1 * (\underline{I}-\underline{L}) + n_3\underline{H}_1 * (\underline{I}-\underline{L})^3$$

and substitution of the series $\underline{L} = \underline{L}_1 + \underline{L}_2 + \underline{L}_3 + \ldots$ in this expression yields

$$\underline{L}_1 + \underline{L}_2 + \underline{L}_3 + \ldots = \underline{H}_1 * (\underline{I}-\underline{L}_1-\underline{L}_2-\underline{L}_3-\ldots) + n_3\underline{H}_1 * (\underline{I}-\underline{L}_1-\underline{L}_2-\underline{L}_3-\ldots)^3$$

$$= (\underline{H}_1*(\underline{I}-\underline{L}_1)-\underline{H}_1*\underline{L}_2-\underline{H}_1*\underline{L}_3-\ldots)$$

$$+ \left(n_3\underline{H}_1*(\underline{I}-\underline{L}_1)^3+3n_3\underline{H}_1*\left((I-L)^2\cdot\underline{L}_2\right)+\ldots\right) \tag{85}$$

Therefore

$$\underline{L}_1 = \underline{H}_1 * (\underline{I}-\underline{L}_1) \tag{86}$$

$$\underline{L}_2 = -\underline{H}_1 * \underline{L}_2 \tag{87}$$

$$\underline{L}_3 = -\underline{H}_1 * \underline{L}_3 + n_3\underline{H}_1 * (\underline{I}-\underline{L})^3 \tag{88}$$

Rearranging Eq. 86 (in a manner similar to the rearrangement that gave Eq. 79 from Eq. 72) yields

$$\underline{L}_1 = (\underline{I}+\underline{H}_1)^{-1} * \underline{H}_1$$

Equation 87 is satisfied for $\underline{L}_2 = 0$, and this is the only solution (see sec. 6.3). Rearrangement of Eq. 88 gives

$$\underline{L}_3 = n_3(\underline{I}+\underline{H}_1)^{-1} * \underline{H}_1 * (\underline{I}-\underline{L}_1)^3 = n_3\underline{L}_1 * (\underline{I}-\underline{L}_1)^3$$

Continuing this procedure gives $\underline{L}_4$, $\underline{L}_5$, and so on. In particular, it can be shown that

$$\underline{L}_4 = 0$$

$$\underline{L}_5 = 3n_3\underline{L}_1 * \left((\underline{I}-\underline{L}_1)^2\cdot\underline{L}_3\right)$$

$$\underline{L}_6 = 0$$

$$\underline{L}_7 = n_3\underline{L}_1 * \left(3(\underline{I}-\underline{L}_1)\cdot\underline{L}_3^2\right) + \left((\underline{I}-\underline{L}_1)^2\cdot\underline{L}_5\right)$$

2.9 IMPULSE RESPONSES AND TRANSFORMS

It has been shown how the algebra of systems can be used to combine systems. But before the output of a system so described can be obtained for some given input, this algebra must be related to the system impulse responses and transforms. We shall give the relation between the algebraic terms and the corresponding impulse responses and transforms.

By means of this algebra, a system $\underline{L}$ is expanded in a series $\underline{L} = \underline{L}_1 + \underline{L}_2 + \ldots + \underline{L}_n + \ldots$, where the $\underline{L}_n$ are given in terms of the system's component subsystems. For an n^{th}-order term of the form $\underline{L}_n = \underline{A}_n + \underline{B}_n$ or $\underline{L}_n[x] = \underline{A}_n[x] + \underline{B}_n[x]$, the corresponding functional equation is

$$\int \cdots \int l_n(\tau_1, \ldots, \tau_n)\, x(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n$$

$$= \int \cdots \int a_n(\tau_1, \ldots, \tau_n)\, x(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n$$

$$+ \int \cdots \int b_n(\tau_1, \ldots, \tau_n)\, x(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n$$

$$= \int \cdots \int \{a_n(\tau_1, \ldots, \tau_n)+b_n(\tau_1, \ldots, \tau_n)\}\, x(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n$$

Therefore

$$l_n(\tau_1, \ldots, \tau_n) = a_n(\tau_1, \ldots, \tau_n) + b_n(\tau_1, \ldots, \tau_n)$$

Hence, for the algebraic term $\underline{L}_n$, where $\underline{L}_n = \underline{A}_n + \underline{B}_n$, the corresponding impulse response is

$$l_n(t_1, \ldots, t_n) = a_n(t_1, \ldots, t_n) + b_n(t_1, \ldots, t_n)$$

The corresponding transform relation is

$$L_n(s_1, \ldots, s_n) = A_n(s_1, \ldots, s_n) + B_n(s_1, \ldots, s_n)$$

Similarly, it can be shown that for the simple multiplication combination, with $\underline{L}_{n+m} = \underline{A}_n \cdot \underline{B}_m$, the corresponding impulse response is

$$l_{n+m}(t_1, \ldots, t_{n+m}) = a_n(t_1, \ldots, t_n)\, b_m(t_{n+1}, \ldots, t_{n+m}) \tag{89}$$

The corresponding transform is

$$L_{n+m}(s_1, \ldots, s_{n+m}) = A_n(s_1, \ldots, s_n)\, B_m(s_{n+1}, \ldots, s_{n+m}) \tag{90}$$

For the cascade situation, with $\underline{L}_n = \underline{A}_1 * \underline{B}_n$, the impulse response is

$$l_n(t_1, \ldots, t_n) = \int a_1(\tau)\, b_n(t_1-\tau, t_2-\tau, \ldots, t_n-\tau)\, d\tau \tag{91}$$

and the transform is

$$L_n(s_1, \ldots, s_n) = A_1(s_1+s_2+\ldots+s_n)\, B_n(s_1, \ldots, s_n) \tag{92}$$

The more general cascade operation also has a relation with a corresponding impulse response and transform. If

$$\underline{L}_{p+g+\ldots+r} = \underline{A}_n \circ (\underline{B}_p \cdot \underline{C}_q \cdot \ldots \cdot \underline{P}_r) \tag{93}$$

then

$$l_{p+q+\ldots+r}(t_1, \ldots, t_{p+q+\ldots+r}) = \int \cdots \int a_n(\tau_1, \ldots, \tau_n)\, b_p(t_1-\tau_1, \ldots, t_p-\tau_1)$$

$$\times c_q(t_{p+1}-\tau_2, \ldots, t_{p+q}-\tau_2) \ldots d\tau_1 \ldots d\tau_n \tag{94}$$

and

$$L_{p+q+\ldots+r}(s_1, \ldots, s_{p+q+\ldots+r}) = A_n(s_1+\ldots+s_p, s_{p+1}+\ldots+s_{p+q}, \ldots)$$

$$\times B_p(s_1, \ldots, s_p)\, C_q(s_{p+1}, \ldots, s_{p+q}) \ldots \tag{95}$$

Some of these combined forms, as written, are not symmetrical, but they can be symmetrized, if it is desired. As we have stated, the impulse response $h_2(t_1, t_2)$ can be symmetrized by forming

$$\frac{h_2(t_1, t_2) + h_2(t_2, t_1)}{2} \tag{96}$$

The transform $H_2(s_1, s_2)$ can be symmetrized by forming

$$\frac{H_2(s_1, s_2) + H_2(s_2, s_1)}{2} \tag{97}$$

Similarly, for $H_3(s_1, s_2, s_3)$, we can construct

$$\frac{1}{6}\{H_3(s_1, s_2, s_3)+H_3(s_1, s_3, s_2)+H_3(s_2, s_2, s_1)+H_3(s_2, s_3, s_1)$$

$$+H_3(s_3, s_1, s_2)+H_3(s_3, s_2, s_1)\} \tag{98}$$

In general, for $H_n(s_1, \ldots, s_n)$, we add up the H_n with all possible arrangements of $s_1, \ldots, s_n$ and divide by the number of arrangements.

Two examples of obtaining the transforms from this algebra will be given. For the feed-through system $\underline{L}$ (see Sec. VI):

$$\underline{L} = \underline{L}_1 + \underline{L}_3 \tag{99}$$

$$\underline{L}_1 = n_1\underline{A}_1 * \underline{B}_1 \tag{100}$$

$$\underline{L}_3 = n_3\underline{A}_1 * \underline{B}_1^3 \tag{101}$$

Let $\underline{A}_1$ have a transform, $\underline{A}_1(s)$, and $\underline{B}_1$ have a transform, $B_1(s)$. We want to find $L_1(s)$, the transform of $\underline{L}_1$, and $L_3(s_1, s_2, s_3)$, the transform of $\underline{L}_3$. By application of Eqs. 89-98, we have $L_1(s) = n_1A_1(s)\, B_1(s)$. From Eq. 90, $\underline{B}_1^2$ has a transform, $B_1(s_1)\, B_1(s_2)$, and $\underline{B}_1^3 = \underline{B}_1^2 \cdot \underline{B}_1$ has a transform, $B_1(s_1)\, B_1(s_2)\, B_1(s)$. Equation 92 then shows that

$$L_3(s_1, s_2, s_3) = n_3A_1(s_1+s_2+s_3)\, B_1(s_1)\, B_1(s_2)\, B_1(s_3)$$

The second system is an example of a feedback system (see sec. 2.8), with

$$\underline{L}_1 = (\underline{I}+\underline{H}_1)^{-1} * \underline{H}_1 \tag{102}$$

$$\underline{L}_3 = n_3\underline{L}_1 * (\underline{I}-\underline{L}_1)^3 \tag{103}$$

$$\underline{L}_5 = 3n_3\underline{L}_1 * \left((\underline{I}-\underline{L}_1)^2 \cdot \underline{L}_3\right) \tag{104}$$

Let $\underline{H}_1$ have a transform, $H_1(s) = A/(s+\alpha)$, where $A \gg \alpha$. Then $(\underline{I}+\underline{H}_1)$ has a transform

$$1 + \frac{A}{s+\alpha} \approx \frac{s+A}{s+\alpha}$$

and, from linear theory, we know that $(\underline{I}+\underline{H}_1)^{-1}$ has a transform

$$\frac{1}{1+H_1(s)} \approx \frac{s+\alpha}{s+A}$$

Then, from Eq. 92, $\underline{L}_1$ has a transform

$$L_1(s) \approx \frac{s+\alpha}{s+A}\,\frac{A}{s+\alpha} = \frac{A}{s+A} \tag{105}$$

Since $(\underline{I}-\underline{L}_1)$ has a transform $I - L_1(s) \approx s/(s+A)$, $(\underline{I}-\underline{L}_1)^2$ has a transform

$$\frac{s_1}{s_1+A}\,\frac{s_2}{s_2+A}$$

from Eq. 90, and $(\underline{I}-\underline{L}_1)^3$ has a transform

$$\frac{s_1}{s_1+A}\,\frac{s_2}{s_2+A}\,\frac{s_3}{s_3+A}$$

Therefore, application of Eq. 92 to Eq. 103 shows that $\underline{L}_3$ has a transform

$$L_3(s_1, s_2, s_3) \approx \frac{n_3 A}{s_1+s_2+s_3+A}\,\frac{s_1}{s_1+A}\,\frac{s_2}{s_2+A}\,\frac{s_3}{s_3+A} \tag{106}$$

Also, since $(\underline{I}-\underline{L}_1)^2 \cdot \underline{L}_3$ has a transform

$$\frac{s_1}{s_1+A}\,\frac{s_2}{s_2+A}\,L_3(s_3, s_4, s_5)$$

$\underline{L}_5$ (Eq. 104) has a transform

$$L_5(s_1, \ldots, s_5) \approx \frac{3n_3 A}{s_1+\ldots+s_5+A}\,\frac{s_1}{s_1+A}\,\frac{s_2}{s_2+A}\,L_3(s_3, s_4, s_5) \tag{107}$$

With some experience the transforms can be readily obtained by inspection from the algebraic equations. We are still not in a position to use these transforms to compute the system output for a given input. However, at the end of Section III, these transforms will be used for this purpose.

2.10 SUMMARY

We have been concerned with expressing nonlinear systems in terms of their linear subsystems and nonlinear no-memory subsystems. The main tool for combining systems has been an algebra of systems. The algebraic manipulations required for system combination obey laws similar to those of other algebras. If the algebra of systems were not used, system combination would have to proceed with involved formulas and by a series of clumsy steps. Our algebraic notation consists of a system representation in which only those aspects of the functional representation that are involved in system combination are emphasized. This algebra applies the powerful concepts of operator mathematics to nonlinear systems.

The relation between the algebraic representation and the system impulse responses and transforms has been shown. Particular emphasis has been placed on the transforms in the two examples presented.

III. SYSTEM TRANSFORMS

3.1 INTRODUCTION

We have represented a nonlinear system in terms of its impulse responses $h_n(t_1, \ldots, t_n)$, or the transforms $H_n(s_1, \ldots, s_n)$. The system output, f(t), is given by Eqs. 2 and 3. The problem, now, is to obtain the $f_n(t)$, and thereby the system output, f(t).

In Section I multidimensional transforms were introduced, and we found that the value of these transforms – just as in the linear case – lies in their making it possible

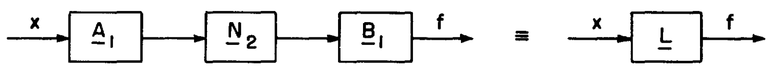


Fig. 16. Illustrative feed-through system.

to substitute multiplications for convolutions. Not only is this true in calculating the system output, but also in cascading systems. This is shown by Eqs. 91 and 92, and by Eqs. 94 and 95.

Another reason for using transforms is that the form of the impulse responses, even for simple systems, is rather complicated. For example, consider the system of Fig. 16. In this case,

$$\underline{L} = \underline{L}_2 = \underline{B}_1 * \underline{N}_2 * \underline{A}_1$$

and $\underline{A}_1$ has a transform $A/(s+\alpha)$, $\underline{B}_1$ has a transform $B/(s+\beta)$, and $n_2 = 1$. Therefore, from Eqs. 90 and 92, $\underline{L}_2$ has a transform

$$L_2(s_1, s_2) = \frac{A^2B}{(s_1+s_2+\beta)(s_1+\alpha)(s_2+\alpha)} \tag{108}$$

Reference to Eqs. 89 and 91 shows that the impulse response is

$$l_2(t_1, t_2) = \int_0^{t_1 \text{ or } t_2} Be^{-\beta\tau}\, A^2 e^{-\alpha(t_1-\tau)}\, e^{-\alpha(t_2-\tau)}\, d\tau$$

for t_1, $t_2 \geq 0$, since $A/(s+\alpha)$ has an inverse, $A \exp(-\alpha t)$, and $B/(s+\beta)$ has an inverse, $B \exp(-\beta t)$. The form of the limit follows because $\underline{A}_1$ and $\underline{B}_1$ are realizable systems, and τ is integrated from 0 to t_1 or t_2, whichever is smaller. Working out the integral gives

$$l_2(t_1, t_2) = \left(\frac{BA^2}{\beta-2\alpha}\right)\left\{e^{-\alpha t_1}\, e^{-\alpha t_2} - e^{-(\beta-\alpha)t_1}\, e^{-\alpha t_2}\right\}$$

for t_1, $t_2 \geq 0$ and $t_1 < t_2$, and

$$l_2(t_1, t_2) = \left(\frac{BA^2}{\beta-2\alpha}\right)\left\{e^{-\alpha t_1} e^{-\alpha t_2} - e^{-\alpha t_1} e^{-(\beta-\alpha)t_2}\right\} \tag{109}$$

for $t_1, t_2 \geq 0$ and $t_2 < t_1$. Comparing this result with Eq. 108 shows the simplicity of the transform, as compared with the impulse response.

Our object, now, is to show how the transforms can be used to determine the output of a system. Emphasis will be placed on an important special case for which the transforms are factorizable. This situation arises when a nonlinear system is lumped.

We shall be in a position to apply the functional representation to the solution of nonlinear system problems, and several examples will be given.

3.2 MULTIDIMENSIONAL TRANSFORMS

Higher-order transforms were defined by Eqs. 10 and 11, and a method of using the transforms was indicated. The linear case is well known. If

$$f_1(t) = \int h_1(\tau)\, x(t-\tau)\, d\tau \tag{110}$$

then

$$F_1(s) = H_1(s)\, X(s) \tag{111}$$

Consider the quadratic system

$$f_2(t) = \iint h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \tag{112}$$

To use transform theory here, we must artificially introduce a t_1 and a t_2, so that

$$f_{(2)}(t_1, t_2) = \iint h_2(\tau_1, \tau_2)\, x(t_1-\tau_1)\, x(t_2-\tau_2)\, d\tau_1 d\tau_2$$

and then

$$F_{(2)}(s_1, s_2) = H_2(s_1, s_2)\, X(s_1)\, X(s_2) \tag{113}$$

Formally, at least, $F_{(2)}(s_1, s_2)$ could be inverted to give $f_{(2)}(t_1, t_2)$, and when $f_2(t)$ is the desired output, $f_2(t) = f_{(2)}(t, t)$. This is illustrated in Fig. 17. We have $f_2(t_1, t_2)$, which could be plotted by contours on the t_1, t_2 plane, but we are only interested in $f_2(t_1, t_2)$ along the 45° line where $t_1 = t_2 = t$. This method generalizes to higher-order cases. For example,

$$f_{(3)}(t_1, t_2, t_3) = \iiint h_3(\tau_1, \tau_2, \tau_3)\, x(t_1-\tau_1)\, x(t_2-\tau_2)\, x(t_3-\tau_3)\, d\tau_1 d\tau_2 d\tau_3$$

but the quantity of interest is $f_3(t)$, with

$$f_3(t) = f_{(3)}(t, t, t) \tag{114}$$

The procedure of taking a number of variables $t_1, \ldots, t_n$ as equal will be called "associating" the variables. The procedure that has been outlined is not particularly

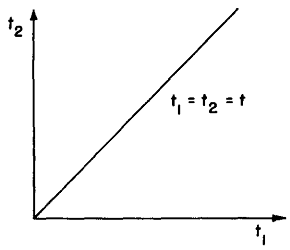


Fig. 17. (t_1, t_2) plane showing $t_1 = t_2$ line.

practical, since it involves taking an n-dimensional inverse transform. A better procedure is to associate the time variables in the transform or frequency domain. That is, given $F_{(2)}(s_1, s_2)$ as the transform of $f_{(2)}(t_1, t_2)$, then $F_2(s)$, the transform of $f_2(t)$, will be found directly from $F_2(s_1, s_2)$. The formal relation is

$$F_2(s) = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} F_{(2)}(s-u, u)\, du \tag{115}$$

where σ is a suitably chosen real number. A proof is given in Appendix A.2. This relation is similar to the Real Multiplication Theorem of linear theory (9). For higher-order transforms, Eq. 115 can be applied successively to associate the variables, two at a time. Then, for example, for $F_{(3)}(s_1, s_2, s_3)$,

$$F_3(s) = \left(\frac{1}{2\pi j}\right)^2 \int_{\sigma-j\infty}^{\sigma+j\infty} \int_{\sigma-j\infty}^{\sigma+j\infty} F_{(2)}(s-u_1, u_1-u_2, u_2)\, du_1 du_2 \tag{116}$$

This is still not very practical because convolutions must be made in the transform domain. The great value of making the associations in the transform domain lies in the fact that these associations can be made by inspection in a large class of problems. This class is the nonlinear generalization of the linear situation in which the transforms are factorizable. The constraint on the system is that it be lumped — that is, that all the transforms of the linear subsystems be factorizable.

Then for the system $\underline{H}$, where $\underline{H} = \underline{H}_1 + \underline{H}_2 + \ldots + \underline{H}_n + \ldots$, we have

$$H_1(s) = \sum_{i=1}^{N} \frac{P_i}{s + p_i} + \sum_{i=0}^{M} R_i s^i \tag{117}$$

where P_i, p_i, and R_i are complex constants. This is familiar from linear theory, and note that terms of the form $P_i/(s+p_i)^n$, for $n > 1$, have been left out. Such terms will be considered separately. If $X(s)$ is the transform of the input to $\underline{H}$, then the transform of the output from the linear portion $\underline{H}_1$ is given by $F_1(s) = H_1(s) X(s)$. If $Y(s)$ is factorizable, then it is known from linear theory that $F_i(s)$ has the same form as Eq. 117, if multiple-order poles are neglected. In the class of systems that is being studied (linear subsystems with memory and nonlinear no-memory subsystems) the most general second-order term is a summation of terms of the form

$$\underline{A}_1 * (\underline{B}_1 \cdot \underline{C}_1) \tag{118}$$

The determination of the transform of such a term was considered in Section II. It is

$$A_1(s_1+s_2)\, B_1(s_1)\, C_1(s_2)$$

where $A_1(s)$, $B_1(s)$, $C_1(s)$ are the transforms of the systems $\underline{A}_1$, $\underline{B}_1$, and $\underline{C}_1$, respectively. If the input has a transform $X(s)$, then the contribution to the system output that is attributable to the output from the term of Eq. 118 has a second-order transform

$$A_1(s_1+s_2)\, B_1(s_1)\, C_1(s_2)\, X(s_1)\, X(s_2) \tag{119}$$

If $B_1(s)$, $C_1(s)$, and $X(s)$ are of the same form as Eq. 117, then $B_1(s_1)\, X(s_1)$ and $C_1(s_2)\, X(s_2)$ have this form, and Eq. 119 becomes

$$A_1(s_1+s_2) \sum_i \frac{B_i}{s_1 + \beta_i} \frac{C_i}{s_2 + \gamma_i} \tag{120}$$

where B_i, C_i, β_i, and γ_i are complex constants. The transform $A_1(s)$ does not have to be factorizable, but it will generally be assumed to be so. Note that the terms $\sum_{i=0}^{M} R_i s^i$ have been excluded from the summation of Eq. 120. This is done because these terms are the transforms of impulses, doublets, and so forth, and such functions do not exist when squared. Should these idealizations occur in a physical problem, they must be removed and replaced by the real physical signals.

The inspection technique can now be developed. Consider a typical term in the second-order case (Eq. 120):

$$G_{(2)}(s_1, s_2) = A_1(s_1+s_2) \frac{B}{s_1 + \beta} \frac{C}{s_2 + \gamma}$$

Application of the transform-domain association equation (Eq. 115) gives

$$G_2(s) = \frac{1}{2\pi j} \int G_{(2)}(s-u, u)\, du \tag{121}$$

or

$$G_2(s) = \frac{1}{2\pi j} \int A_1(s-u+u) \frac{B}{s-u+\beta} \frac{C}{u+\gamma}\, du$$

$$= A_1(s) \frac{1}{2\pi j} \int \frac{B}{s-u+\beta} \frac{C}{u+\gamma}\, du$$

The term that is to be considered is

$$\frac{1}{2\pi j}\int \frac{B}{s-u+\beta}\,\frac{C}{u+\gamma}\,du \tag{122}$$

But $(B/(s_1+\beta)(C/(s_2+\gamma))$ is easily inverted, and has an inverse transform

$$Be^{-\beta t_1}\,Ce^{-\gamma t_2} \qquad t_1, t_2 \geq 0$$

Setting $t_1 = t_2 = t$ gives

$$B\,Ce^{-(\beta+\gamma)t} \qquad t \geq 0$$

This has a transform, $BC/[s+(\beta+\gamma)]$, and it is seen that

$$\frac{1}{2\pi j}\int \frac{B}{s-u+\beta}\,\frac{C}{u+\gamma}\,du = \frac{BC}{s+(\beta+\gamma)} \tag{123}$$

Finally, we have

$$G_2(s) = A_1(s)\,\frac{BC}{s+(\beta+\gamma)}$$

where $G_2(s)$ is the transform of $g_2(t)$; and $g_2(t) = g_{(2)}(t,t)$, where $g_{(2)}(t_1,t_2)$ is the inverse transform of $G_{(2)}(s_1, s_2)$. That is, we have made the association of t_1 and t_2 by a transform-domain manipulation that gives us the ordinary linear transform of the desired time function $g_2(t)$. Furthermore, this manipulation can be done by inspection.

That it is an inspection technique is seen by noting that the association of t_1 and t_2 changes

$$G_{(2)}(s_1, s_2) = A_1(s_1+s_2)\,\frac{B}{s_1+\beta}\,\frac{C}{s_2+\gamma} \tag{124}$$

into

$$G_2(s) = A_1(s)\,\frac{BC}{s+(\beta+\gamma)} \tag{125}$$

Examination of Eqs. 124 and 125 shows that the change is a very obvious one and can be obtained by inspection.

Higher-order transforms can be reduced by applying the inspection procedure to associate the variables, two at a time. For example, consider the third-order term

$$\frac{A}{s_1+s_2+s_3+\alpha}\,\frac{B}{s_2+s_3+\beta}\,\frac{C}{s_1+\gamma}\,\frac{C}{s_2+\gamma}\,\frac{C}{s_3+\gamma} \tag{126}$$

Application of the formal association equation (Eq. 115) to associate s_2 and s_3 yields

$$\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{A}{s_1+s_2-u+u+\alpha}\,\frac{B}{s_2-u+u+\beta}\,\frac{C}{s_1+\gamma}\,\frac{C}{s_2-u+\gamma}\,\frac{C}{u+\gamma}\,du$$

$$=\frac{A}{s_1+s_2+\alpha}\,\frac{B}{s_2+\beta}\,\frac{C}{s_1+\gamma}\,\frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\frac{C}{s_2-u+\gamma}\,\frac{C}{u+\gamma}\,du \tag{127}$$

This integration is of the same type as that in Eq. 122, and it yields $C/(s_2+2\gamma)$ (see Eq. 123). Therefore Eq. 127 becomes

$$\frac{A}{s_1+s_2+\alpha}\,\frac{B}{s_2+\beta}\,\frac{C}{s_1+\gamma}\,\frac{C}{s_2+2\gamma} \tag{128}$$

(For convenience, the procedure of associating two time variables t_i and t_j in the frequency domain will be called "associating" the frequency variables s_i and s_j.) The change from Eq. 126 to Eq. 128 is obtained by applying the inspection technique to the variables s_2 and s_3. Now, Eq. 128 equals

$$\frac{A}{s_1+s_2+\alpha}\,\frac{C}{s_1+\gamma}\,\frac{BC^2}{2\alpha-\beta}\left\{\frac{1}{s_2+\beta}-\frac{1}{s_2+2\gamma}\right\} \tag{129}$$

and the association procedure can be applied to associate s_1 and s_2. The result is

$$\frac{BC^2}{2\alpha-\beta}\,\frac{A}{s+\alpha}\left\{\frac{1}{s+(\beta+\gamma)}-\frac{1}{s+3\gamma}\right\} \tag{130}$$

Similarly, a transform of any order can be reduced to a first-order transform by successive use of the inspection technique. For example, consider the fourth-order term

$$K_1(s_1+s_2+s_3+s_4)\,\frac{A}{s_1+s_2+\alpha}\,\frac{B}{s_1+\beta}\,\frac{B}{s_2+\beta}\,\frac{C}{s_3+\gamma}\,\frac{C}{s_4+\gamma} \tag{131}$$

where $K_1(s)$ is some transform function. Associating s_3 and s_4 by inspection yields

$$K_1(s_1+s_2+s_3)\,\frac{A}{s_1+s_2+\alpha}\,\frac{B}{s_1+\beta}\,\frac{B}{s_2+\beta}\,\frac{C^2}{s_3+2\gamma}$$

Next, associate s_1 and s_2. The result is

$$K_1(s_2+s_3)\,\frac{A}{s_2+\alpha}\,\frac{B}{s_2+2\beta}\,\frac{C^2}{s_3+2\gamma}$$

$$=K_1(s_2+s_3)\,\frac{AB}{2\beta-\alpha}\left(\frac{1}{s_2+\alpha}-\frac{1}{s_2+2\beta}\right)\frac{C^2}{s_3+2\gamma}$$

Finally, s_2 and s_3 can be associated, and we obtain

$$K_1(s)\,\frac{AB}{2\beta-\alpha}\left(\frac{1}{s+\alpha+2\gamma}-\frac{1}{s+2\beta+2\alpha}\right) \tag{132}$$

Notice that s_3 and s_4 were associated first, then s_1 and s_2, and finally s_2 and s_3. If we had associated s_3 and s_4, s_3 and s_2, and s_2 and s_1, we would have had to handle a form that had not been discussed. At times, when we are using the inspection technique, it will be necessary to associate the variables in a definite order to avoid forms that we cannot handle with the method discussed here. In a similar manner, fifth-, sixth-, and higher-order transforms can be reduced to first-order transforms.

The method for using multidimensional transforms can be summarized as follows:

(a) Introduce artificial variables $t_1, t_2, \ldots, t_n$, so that multidimensional transforms can be used to specify the system output.

(b) Associate these variables $t_1, \ldots, t_n$ with the time variable t by means of the inspection procedure in the transform domain. The result of this procedure is the transform of the system output.

(c) Then, if it is desired, this first-order transform can be inverted by the ordinary linear system analysis methods to give $f_1(t)$, $f_2(t)$, and so on, where the output is

$$f(t) = f_1(t) + f_2(t) + \ldots + f_n(t) + \ldots$$

Otherwise, the output signal can be interpreted in the frequency domain, as is often done in linear system analysis.

Nonfactorizable higher-order transforms – for example, situations in which delay is involved – can often be handled by solving the association formula (Eq. 115) in the manner given by Eqs. 121-125, that is, by working partly in the time domain and partly in the frequency domain.

As an example, consider the transform

$$\frac{A}{(s_1+\alpha)^n} \frac{B}{(s_2+\beta)^m} \tag{133}$$

where s_1 and s_2 are to be associated. This is the multiple-pole situation which we have ignored previously (Eq. 120). Equation 133 is easily inverted and has the transform

$$\frac{A}{(n-1)!} t_1^{n-1} e^{-\alpha t_1} \frac{B}{(m-1)!} t_2^{m-1} e^{-\beta t_2}$$

Associating t_1 and t_2 yields

$$\frac{A}{(n-1)!} \frac{B}{(m-1)!} t^{n+m-2} e^{-(\alpha+\beta)t}$$

and this has a transform

$$AB \frac{(n+m-2)!}{(n-1)!\,(m-1)!} \frac{1}{(s+\alpha+\beta)^{n+m-1}} \tag{134}$$

which is the result of associating s_1 and s_2 in Eq. 133.

Before giving some examples of the application of the material already presented, we shall round out the discussion of system transforms by considering some other properties of these higher-order transforms.

3.3 STEADY-STATE RESPONSE

In linear system theory with $f_1 = \underline{H}_1[x]$, and $x(t) = \text{Re}\,\{Xe^{j\omega t}\}$, where X is a complex constant, it is well known that in the steady state, the output $f_1(t)$ is given by

$$f_1(t) = \text{Re}\left\{XH_1(j\omega)\,e^{j\omega t}\right\}$$

where $H_1(j\omega)$ is the system transform $H_1(s)$ evaluated at $s = j\omega$.

A similar result is found for the higher-order system transforms $H_n(s_1, \ldots, s_n)$. To develop the steady-state output of a second-order system with a sinusoidal input, consider the second-order operation on an input pair:

$$g_2 = \underline{H}_2(xy) \tag{135}$$

The complex functions are

$$x(t) = Xe^{j\omega_1 t}$$

and

$$y(t) = Ye^{j\omega_2 t}$$

where X and Y are complex constants. The steady-state value of $g_2(t)$ is given by

$$g_2(t) = XYH_2(j\omega_1, j\omega_2)\,e^{j\omega_1 t}\,e^{j\omega_2 t} \tag{136}$$

where $H_2(j\omega_1, j\omega_2)$ is $H_2(s_1, s_2)$ evaluated at $s_1 = j\omega_1$ and $s_2 = j\omega_2$. We see that the transform $H_2(s_1, s_2)$ has a steady-state interpretation very similar to the linear transform $H_1(s)$. The operation of Eq. 135 does not exist alone. In order to examine the real situation, consider the actual second-order system, with $f_2 = \underline{H}_2[x]$. Let $x = y + z$, with

$$y(t) = \frac{X}{2}\,e^{j\omega t}$$

and

$$z(t) = \frac{\bar{X}}{2}\,e^{-j\omega t}$$

Here, $\bar{X}$ is the conjugate of the complex number X. Then

$$x(t) = \text{Re}\,\{Xe^{j\omega t}\}$$

The problem now is to find the steady-state value of $f_2(t)$. We have

$$f_2 = \underline{H}_2((y+z)^2) = \underline{H}_2(y^2) + 2\underline{H}_2(yz) + \underline{H}_2(z^2)$$

and, by use of Eq. 136,

$$f(t) = \frac{1}{2} \operatorname{Re} \left\{ X^2 H_2(j\omega, j\omega) e^{j2\omega t} + X\bar{X} H_2(j\omega, -j\omega) \right\}$$

Hence, the steady-state response of a quadratic system is composed of a dc term and a double frequency term. This is similar to the effect of a no-memory squaring operation.

In a similar manner, the steady-state response of higher-order systems can be formed. For the third-order case,

$$f(t) = \frac{1}{4} \operatorname{Re} \left\{ X^3 H_3(j\omega, j\omega, j\omega) e^{j3\omega t} + 3X^2 \bar{X} H_3(j\omega, j\omega, -j\omega) e^{j\omega t} \right\} \tag{137}$$

It should be noted that the solution of these equations depends upon $H_n(s_1, \ldots, s_n)$ being symmetric. If the operation of taking the real part is omitted, then the quantities $X^3 H_3(j\omega, j\omega, j\omega)$, and so on, can be regarded as complex amplitudes of the corresponding sinusoids, just as in linear system analysis.

Not only do these results furnish an interpretation of the higher-order transforms; they also show that the steady-state response of a system can be easily obtained, once the system transforms are known. To give an example, consider the nonlinear amplifier of Fig. 13. We shall use the system transforms for $L_1(s)$ and $L_3(s_1, s_2, s_3)$ developed in section 2.9.

Let

$$A_1(s) = \frac{As}{(s+\alpha)^2 + \omega_o^2}$$

and

$$B_1(s) = \frac{Bs}{(s+\beta)^2 + \omega_o^2}$$

Then

$$L_1(s) = \frac{n_1 ABs^2}{\left[(s+\alpha)^2 + \omega_o^2\right]\left[(s+\beta)^2 + \omega_o^2\right]}$$

and

$$L_3(s_1, s_2, s_3) = \frac{n_3 A}{\left[(s_1+s_2+s_3+\alpha)^2 + \omega_o^2\right]} \frac{Bs_1}{\left[(s_1+\beta)^2 + \omega_o^2\right]} \frac{Bs_2}{\left[(s_2+\beta)^2 + \omega_o^2\right]} \frac{Bs_3}{\left[(s_3+\beta)^2 + \omega_o^2\right]}$$

If we apply the methods that have been given for obtaining the steady-state sinusoidal response (in particular, Eq. 137), at frequency ω, we have the following complex quantities:

(a) Linear gain,

$$L_1(j\omega) = \frac{n_1 ABX(j\omega)^2}{\left[(j\omega+\beta)^2 + \omega_o^2\right]\left[(j\omega+\alpha)^2 + \omega_o^2\right]} \tag{138}$$

(b) First-harmonic distortion, which is the amplitude of the first-harmonic term that varies as X^3,

$$= \frac{3}{4} L_3(j\omega, j\omega, -j\omega)$$

$$= \frac{3}{4} \frac{n_3 A B^3 X^3 |j\omega|^2 (j\omega)}{\left[(j\omega+a)^2+\omega_o^2\right]} \frac{1}{|(j\omega+\beta)^2+\omega_o^2|^2 \left[(j\omega+\beta)^2+\omega_o^2\right]} \tag{139}$$

(c) Third-harmonic distortion, which is the amplitude of the signal at three times the input frequency,

$$= \frac{1}{4} L_3(j\omega, j\omega, j\omega)$$

$$= \frac{1}{4} \frac{n_3 A B^3 X^3 (j\omega)^3}{\left[(j3\omega+a)^2+\omega_o^2\right]\left[(j\omega+\beta)^2+\omega_o^2\right]^3} \tag{140}$$

where X is the amplitude of the input sinusoid.

3.4 INITIAL-VALUE AND FINAL-VALUE THEOREMS

Another useful property of the higher-order transforms is that they obey initial-value and final-value theorems that are similar to the linear transforms. If $f_{(n)}(t_1, \ldots, t_n$ has a transform $F_{(n)}(s_1, \ldots, s_n)$, and if $f_n(t) = f_{(n)}(t, t, \ldots, t)$, then the following relations are true:

$$\lim_{t \to 0} f_n(t) = \lim_{\substack{s_1 \to \infty \\ \vdots \\ s_n \to \infty}} F_{(n)}(s_1, \ldots, s_n)\, s_1 \ldots s_n \tag{141}$$

and

$$\lim_{t \to \infty} f_n(t) = \lim_{\substack{s_1 \to 0 \\ \vdots \\ s_n \to 0}} F_{(n)}(s_1, \ldots, s_n)\, s_1 \ldots s_n \tag{142}$$

Proofs of Eqs. 141 and 142 are given in Appendix A.3. The usual linear theory constraints hold: all limits, in both the time and frequency domains, must exist.

These results can be used, just as in linear system analysis, to obtain the initial and final values of system output values, slopes, and so forth, rapidly.

3.5 EXAMPLE 2.

This example is concerned with the feedback servo system of Fig. 15. $\underline{H}_1$ is the cascade combination of an armature-controlled dc motor and a gain factor, and

$H_1(s) = A/(s+a)$. The output f is the motor velocity, and $\underline{N}$ is a compensation device. (See Fig. 18.) The objective of this design is to reduce the step response time of the system.

First, consider the linear uncompensated system with $\underline{N} = \underline{I}$. The step response of this system is

$$f(t) = X(1-e^{-At}) \qquad t \geq 0$$

where X is the amplitude of the input step, and $A \gg a$. The rise time of the system can be reduced by increasing the gain factor A, but there is an acceleration constraint that limits the size of A. This limit on A is determined by X_m, the maximum input amplitude with which the system is to be used, and by M, the maximum allowable acceleration. In fact, the maximum gain for this linear system A_ℓ is given by $A_\ell = M/X_m$.

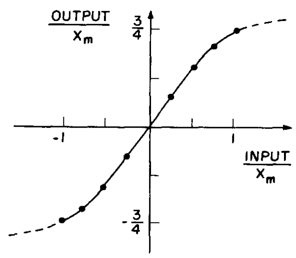


Fig. 18. Example 2. Characterization of $\underline{N}$.

In this problem, we shall show that a simple nonlinear no-memory compensating device, $\underline{N} = \underline{I} + \underline{N}_3$, can be used to decrease the response time and still meet the acceleration constraint. Only the first two terms of the output are significant in this problem, and hence $f(t) = f_1(t) + f_3(t)$. The nonlinear system in this problem is the same as that of Fig. 15, and the first two system transforms have been given in Eqs. 105 and 106. If the input x(t) has a transform X/s, then the output transforms are

$$F_1(s) = \frac{AX}{s(s+A)} \tag{143}$$

and

$$F_{(3)}(s_1, s_2, s_3) = \frac{n_3AX^3}{(s_1+s_2+s_3+A)(s_1+A)(s_2+A)(s_3+A)} \tag{144}$$

By using the inspection technique, we obtain

$$F_3(s) = \frac{n_3AX^3}{(s+A)(s+3A)} \tag{145}$$

and thus

$$f(t) = X\left\{1 - \left(1 - \frac{n_3X^2}{2}\right)e^{-At} - \frac{n_3X^2}{2}e^{-3At}\right\} \tag{146}$$

Also, we have

$$f'(t) = AX\left\{\left(1 - \frac{n_3X^2}{2}\right)e^{-At} + \frac{3}{2}n_3X^2\ e^{-3At}\right\} \tag{147}$$

where f'(t) represents the acceleration. It is possible to investigate various choices of A and n_3 to obtain a rapid response and still have $f'_{max} = M$. A good choice is

$$n_3 = -\frac{1}{4X_m^2} \tag{148}$$

in which case the gain can be taken as

$$A_n = \frac{4}{3}\frac{M}{X_m} \tag{149}$$

and the acceleration constraint is satisfied for the maximum input amplitude, X_m. The 0 to 90 per cent rise time, t_r, for maximum input signal is

$$t_r = 1.8\frac{X_m}{M} \tag{150}$$

and for the uncompensated linear case, it is

$$t_r = 2.3\frac{X_m}{M}$$

Therefore, the rise time can be decreased 20 per cent by the use of simple nonlinear compensation. For small signals, the rise time has been decreased 25 per cent.

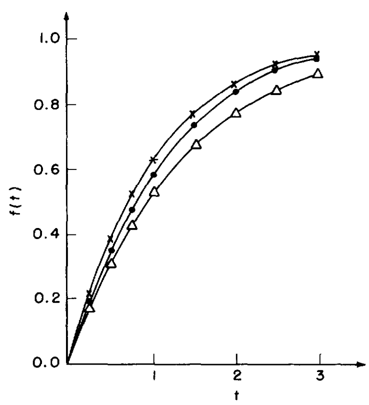


Fig. 19. System response. (All outputs are normalized to 1.) Large-signal input: o, compensated nonlinear system; △, uncompensated linear system. Small-signal input: ×, compensated nonlinear system; △, uncompensated linear system.

Figure 19 shows the transient responses for maximum input steps and very small input steps for the linear uncompensated and the nonlinear compensated systems. In both curves, the output is normalized to 1. Figure 18 gives the input-output characteristic of the nonlinear device. It is specified only for an input less than X_m. Outside this region, any saturation characteristic suffices.

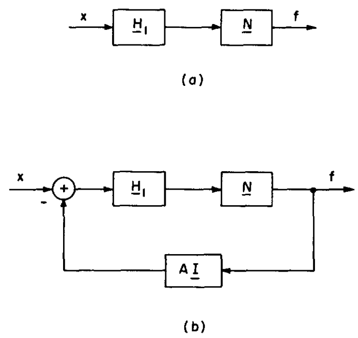


Fig. 20. Example 3. (a) Lowpass amplifier with output distortion. (b) Amplifier A with feedback.

It is appropriate to emphasize the importance of signal amplitude in the analysis and synthesis of nonlinear systems. In the analysis of linear systems, the input-signal amplitude is rather incidental. This is not the case with nonlinear systems because the nature of the system response is greatly dependent upon input amplitude. Therefore in a nonlinear system problem the range of input amplitude is a very important parameter. A knowledge of this range is essential in using the functional representation for system analysis because this will determine how many terms of the output must be retained.

The use of nonlinear compensation in servo systems is a problem of considerable interest. This particular example has been given not only to illustrate the use of the functional representation for nonlinear feedback systems, but also to indicate the possible use of the representation in the study of the general problem of continuous nonlinear compensation.

3.6 EXAMPLE 3.

The systems of Fig. 20 are: A, an amplifier with output distortion, and B, the same amplifier with some weak feedback for reducing distortion. In this situation, the classical steady-state methods do not suffice.

Let $\underline{H}_1$ have a transform, $H/s + a$, and $\underline{N} = \underline{I} + \underline{N}_3 + \underline{N}_5$. The transforms of systems A and B can both be computed by the methods previously explained and illustrated. If the input is $x(t) = \text{Re}\,\{Xe^{j\omega t}\}$, The transforms can be used to give the distortion ratios for the systems. (Transforms and details are given in Appendix B.1.) For low frequencies, these ratios for system A are:

$$\text{First-harmonic distortion} = \frac{3}{4} n_3 B^2 X^2 + \frac{5}{8} n_5 B^4 X^4 \tag{151}$$

$$\text{Third-harmonic distortion} = \frac{1}{4} n_3 B^2 X^2 + \frac{5}{16} n_5 B^4 X^4 \tag{152}$$

$$\text{Fifth-harmonic distortion} = \frac{1}{16} n_5 B^4 X^4 \tag{153}$$

where X is the input amplitude, and $B = HX/a$ is the linear low-frequency gain. Assume that $\frac{3}{4} n_3 B^2 X^2$ is approximately $\frac{1}{5}$, and $\frac{5}{8} n_5 B^4 X^4$ is approximately $\frac{1}{20}$, at the maximum value of the input amplitude, X. Then the distortion ratios for feedback system B are as given below. (See also Appendix B.1.) G is the ratio

$$\frac{H' - H}{H} \tag{154}$$

where the gain factor H has been increased to H' to keep the linear gain of feedback system B equal to that of system A. These ratios are:

$$\text{First-harmonic distortion} \approx \frac{3}{4} n_3 B^2 X^2 + \frac{5}{8} n_5 B^4 X^4 - \frac{15}{8} n_3^2 B^4 X^4 G \tag{155}$$

$$\text{Third-harmonic distortion} \approx \frac{1}{4} n_3 B^2 X^2 + \frac{5}{16} n_5 B^4 X^4 - \frac{15}{16} n_3^2 B^4 X^4 G \tag{156}$$

$$\text{Fifth-harmonic distortion} \approx \frac{1}{16} n_5 B^4 X^4 - \frac{3}{16} n_3^2 B^4 X^4 G \tag{157}$$

We see that feedback can be used to decrease the amount of distortion even with the linear gain kept the same. It is interesting to note that if $n_5 = 3n_3^2 G$, then the distortion from the fifth-order nonlinearities will be completely removed by the feedback.

This example could be extended to higher distortion and stronger feedback by developing more of the terms in the expansion of the feedback system.

3.7 EXAMPLE 4.

The system of Fig. 21a is an example of an FM detector of the phase-locked-loop type. The input is

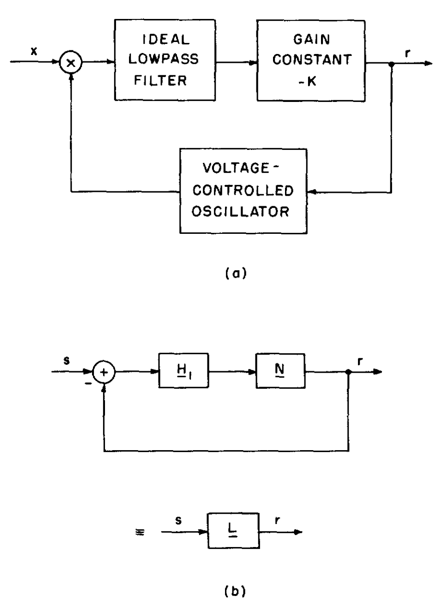


Fig. 21. Example 4. (a) Phase-locked loop. (b) Equivalent system.

$$x(t) = X \cos\left(\omega_o t + \int_{-\infty}^{t} s(\tau)\, d\tau\right) \tag{158}$$

where ω_o is the frequency of the system's voltage-controlled oscillator, X is the signal amplitude, and $s(\tau)$ is the modulating signal. The equation for this multiplicative feedback system is

$$r(t) = -XK\underline{L}_1\left[\cos\left(\omega_o t + \int_{-\infty}^{t} s(\tau)\, d\tau\right) \sin\left(\omega_o t + \int_{-\infty}^{t} r(\tau)\, d\tau\right)\right] \tag{159}$$

where $\underline{L}_1$ is the ideal lowpass filter.

Expanding Eq. 159, we obtain

$$r(t) = -XK\underline{L}_1\left[\sin\left(2\omega_o t + \int_{-\infty}^{t} \{r(\tau)+s(\tau)\}\, d\tau\right) + \sin\left(\int_{-\infty}^{t} \{r(\tau)-s(\tau)\}\, d\tau\right)\right] \tag{160}$$

Since $\underline{L}_1$ is lowpass, the term with frequency centered at $2\omega_o$ can be neglected, and $r = A \sin\{\underline{H}_1[s-r]\}$. $\underline{H}_1$ is an ideal integrator and A is a gain constant, where $A = XK$. A diagram of this equivalent system is shown in Fig. 21b, in which

$$\underline{N}[y] = A \sin y \tag{161}$$

Solving for the first three terms of system $\underline{L}$, we obtain

$$L_1(s) = \frac{A}{s + A} \tag{162}$$

$$L_3(s_1, s_2, s_3) = -\frac{1}{3} \frac{A(s_1+s_2+s_3)}{s_1 + s_2 + s_3 + A} \frac{1}{s_1 + A} \frac{1}{s_2 + A} \frac{1}{s_3 + A} \tag{163}$$

$$L_5(s_1, \dots, s_5) = -\frac{A(s_1+\dots+s_5)}{s_1 + \dots + s_5 + A}\left\{\frac{1}{2} \frac{A}{(s_1+s_2+A)} - \frac{1}{5}\right\} \times \frac{1}{s_1 + A} \cdots \frac{1}{s_5 + A} \tag{164}$$

First, the system step response will be computed. If the input s(t) is a step response Su(t), then a good approximation to the output r(t) for $S^2/A^2 < 0.5$ is given by

$$r(t) = r_1(t) + r_3(t) + r_5(t)$$

Associated with r_1, r_3, and r_5 are the multiple-order transforms:

$$R_1(s) = L_1(s)\,\frac{S}{s}$$

$$R_{(3)}(s_1, s_2, s_3) = L_3(s_1, s_2, s_3)\,\frac{S}{s_1}\,\frac{S}{s_2}\,\frac{S}{s_3}$$

$$R_{(5)}(s_1, \ldots, s_5) = L_5(s_1, \ldots, s_5)\,\frac{S}{s_1}\cdots\frac{S}{s_5}$$

Converting these to the first-order transforms $R_1(s)$, $R_3(s)$, and $R_{(5)}$, and inverting (see Appendix B.2 for details) gives:

$$\begin{aligned} r(t) = S\Bigg\{1 &- \left(1 - \frac{5}{12}\frac{S^2}{A^2} + \frac{65}{48}\frac{S^4}{A^4}\right)e^{-At} \\ &+ \left(\frac{2}{3}\frac{S^2}{A^2} + 5\frac{S^4}{A^4}\right)e^{-2At} - \left(\frac{1}{4}\frac{S^2}{A^2} + 6\frac{S^4}{A^4}\right)e^{-3At} \\ &+ \left(\frac{8}{3}\frac{S^4}{A^4}\right)e^{-4At} - \left(\frac{5}{16}\frac{S^4}{A^4}\right)e^{-5At} \\ &+ \left(\frac{1}{6}\frac{S^2}{A^2} + \frac{5}{4}\frac{S^4}{A^4}\right)At\,e^{-At}\Bigg\} \end{aligned}$$

For small S/A, the system is linear with a response, $r(t) = S(1-e^{-At})$, and it departs significantly from this linear operation as S^2/A^2 approaches 0.5. It should also be noted that if S > A, then the system becomes unstable because the form of $\underline{N}$ (see Eq. 161) restricts the output r to be less than A, and static balance is no longer possible.

The system steady-state distortion with sinusoidal input can be readily obtained by the appropriate substitution of $j\omega$ in the system transforms (Eqs. 162, 163, and 164).

3.8 SUMMARY

The basic material for the analysis of continuous nonlinear systems with deterministic inputs has now been presented. An algebra of systems has been used to describe a system in terms of its component subsystems. From this description the system transforms can be found. These transforms can then be used to determine the system's response to various inputs.

IV. REMARK ON APPLICATIONS OF THE ALGEBRA OF SYSTEMS AND SYSTEM TRANSFORMS

We shall be concerned here with several topics that are extensions of the material presented in Sections I-III. The first topic concerns the use of the algebra of systems for block-diagram manipulations.

4.1 BLOCK-DIAGRAM MANIPULATIONS

An example will be given to illustrate how this algebra can be used to perform block-diagram manipulations. It will be shown how such manipulations can be performed algebraically, rather than through a sequence of diagrams.

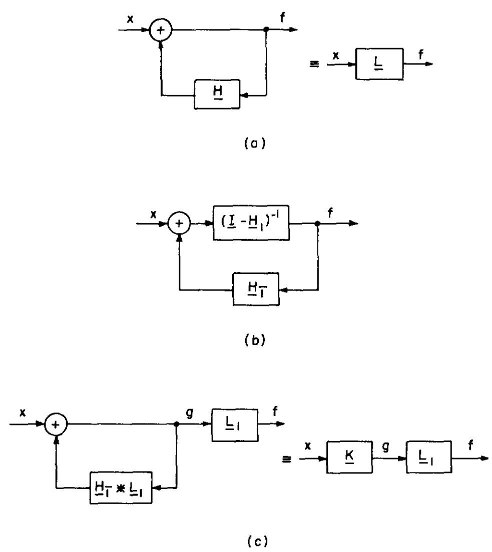


Fig. 22. Block-diagram manipulation: (a) feedback system; (b) first equivalent system; (c) second equivalent system.

Consider the feedback system of Fig. 22a, in which $\underline{H} = \underline{H}_1 + \underline{H}_{\bar{1}}$. $\underline{H}_1$ is the linear part of the system $\underline{H}$, and $\underline{H}_{\bar{1}}$ is the nonlinear part. The object of this example is to show how the linear part of a feedback system can be isolated. We have $\underline{L} = \underline{I} + \underline{H} * \underline{L} = \underline{I} + (\underline{H}_1 + \underline{H}_{\bar{1}}) * \underline{L}$, and then $\underline{L} = \underline{I} + \underline{H}_1 * \underline{L} + \underline{H}_{\bar{1}} * \underline{L}$, or, if we take $\underline{H}_1 * \underline{L}$ over to the left-hand side, we have

$$(\underline{I}-\underline{H}_1) * \underline{L} = \underline{I} + \underline{H}_{\bar{1}} * \underline{L} \tag{165}$$

Then

$$(\underline{I}-\underline{H}_1)^{-1} * (\underline{I}-\underline{H}_1) * \underline{L} = (\underline{I}-\underline{H}_1)^{-1} * (\underline{I}+\underline{H}_{\bar{1}}*\underline{L}) \tag{166}$$

or

$$\underline{L} = (\underline{I}-\underline{H}_1)^{-1} * (\underline{I}+\underline{H}_{\overline{1}}*\underline{L}) \tag{167}$$

Equation 167 is the system equation for the system of Fig. 22b, with the linear part of the system concentrated in the forward loop. Note that the linear part of $\underline{L}$ is given as

$$\underline{L}_1 = (\underline{I}-\underline{H}_1)^{-1}$$

Now, if $\underline{K} = \underline{L}_1^{-1} * \underline{L}$, then $\underline{L} = \underline{L}_1 * \underline{K}$, and from Eq. 165,

$$\underline{K} = \underline{I} + (\underline{H}_{\overline{1}}*\underline{L}_1) * \underline{K} \tag{168}$$

Thus, another equivalent configuration is obtained as shown in Fig. 22c. A third equivalent configuration could also be obtained with $\underline{L}_1$ in front of the nonlinear feedback system.

There are several reasons why such changes in a feedback system may be desired. For example, it might be more desirable to construct the system in one configuration than in another. Or, some particular configuration could be the basis for an alternative system expansion. For example, Zames (10) has developed the concept of expanding a feedback system in a series about the linear part.

4.2 COMPLEX TRANSLATION

The complex translation theorem of the theory of linear analysis (9) can be stated as follows:

If $f(t)$ has a transform $F(s)$, then $e^{-at}f(t)$ has a transform $F(s+a)$.

Here, a is a complex number. A similar theorem holds for higher-order transforms:

If $f_n(t_1,\ldots,t_n)$ has an n-dimensional transform $F_n(s_1,\ldots,s_n)$, then $\exp(-a_1t_1-\ldots-a_nt_n)$ $f(t_1,\ldots,t_n)$ has a transform $F_n(s_1+a_1,\ldots,s_n+a_n)$.

The $a_1, a_2, \ldots, a_n$ are complex numbers, and the proof is essentially the same as the proof for the linear case.

This translation can be useful in finding the envelope response of a system. For a linear system $\underline{H}_1$, with transform $H_1(s)$, let the input be the real part of $x(t)$, where $x(t) = e(t)\exp(j\omega_1 t)$, and $e(t)$ is real. If the complex output, $f(t)$, is in the form

$$f(t) = o(t)\, e^{j\omega_1 t} \tag{169}$$

where $o(t)$ is the complex envelope, then

$$O(s) = H_1(s+j\omega_1)\, E(s) \tag{170}$$

as can be shown by the translation theorem. Then $o(t)$ is the envelope of the output sinusoid.

To illustrate the use of the translation theorem for obtaining the envelope of the output from a higher-order system, consider the third-order system $\underline{H}_3$, with the input

$$z(t) = \text{Re}\left\{x(t)\, e^{j\omega t+\phi}\right\}$$

where x(t) is real. By expanding

$$\underline{H}_3\left(\left(\frac{x(t)}{2}\, e^{j\omega t+\phi} + \frac{x(t)}{2}\, e^{-j\omega t-\phi}\right)^3\right)$$

as we did for the steady-state situation, and applying the translation theorem, it can be shown that the complex envelope of the third-harmonic output has a third-order transform:

$$\frac{1}{4} X(s_1)\, X(s_2)\, X(s_3)\, H_3(s_1+j\omega, s_2+j\omega, s_3+j\omega) \tag{171}$$

when X(s) is the transform of x(t). The third-order transform of the enevlope of the first-order harmonic is

$$\frac{3}{4} X(s_1)\, X(s_2)\, X(s_3)\, H_3(s_1+j\omega, s_2+j\omega, s_3-j\omega) \tag{172}$$

and the associated first-order transforms can then be found by the methods of Section III. This procedure for finding envelope responses generalizes, in a straightforward manner, to systems of any order.

As an example of the calculation of envelope responses of nonlinear systems, consider the feed-through example of section 2.6. For this system, $L_1(s)$ and $L_3(s_1,s_2,s_3)$ were developed in section 3.3.

Assume that α is sufficiently small that the third-harmonic output from the system is negligible, and let the input x(t) be

$$x(t) = X \cos \omega_o t \qquad t \geq 0$$

$$= O \qquad t < 0$$

Then the output can be shown to be

$$f(t) = (o_1(t)+o_3(t)) \cos \omega_o t \qquad t \geq 0$$

The transform of $o_1(t)$ is $O_1(s)$, and a third-order transform, $O_3(s_1, s_2, s_3)$, and a first-order transform, $O_3(s)$, are associated with $o_3(t)$. Thus

$$O_1(s) = L_1(s+j\omega)\frac{X}{s} \approx \frac{n_1ABX}{(s+\alpha)(s+\beta)}$$

and

$$O_3(s_1, s_2, s_3) = L_3(s_1+j\omega, s_2+j\omega, s_3-j\omega)\, \frac{X}{s_1}\, \frac{X}{s_2}\, \frac{X}{s_3}$$

$$\approx \frac{3}{4}\, \frac{n_3AB^3X^3}{(s_1+s_2+s_3+\alpha)(s_1+\beta)(s_2+\beta)(s_3+\beta)}$$

It is assumed that α and β are much less than ω_o, and poles far away from the origin have been neglected. By use of the inspection methods of Section III, $O_3(s)$ is obtained from $O_{(3)}(s_1, s_2, s_3)$, and

$$O_3(s) = \frac{3}{4} \frac{n_3 K H^3 X^3}{(s+\alpha)(s+3\beta)}$$

Inverting $O_1(s)$ and $O_3(s)$ yields

$$o_1(t) = \frac{n_1 HKX}{(\beta-\alpha)} (e^{-\alpha t} - e^{-\beta t}) \qquad \text{for } t \geq 0$$

and

$$o_3(t) = \frac{n_3 K H^3 X^3}{(3\beta-\alpha)} (e^{-\alpha t} - e^{-3\beta t}) \qquad \text{for } t \geq 0$$

where $(o_1(t)+o_3(t))$ is the envelope of the output sinusoid.

4.3 A FINAL-VALUE THEOREM

A variation of the final-value theorem (see sec. 3.4) will now be given. If $y(t_1, \ldots, t_n)$ has a transform $Y(s_1, \ldots, s_n)$, then

$$\lim_{t_1 \to \infty} y(t_1, \ldots, t_n) = \lim_{s_i \to 0} Y(s_1, \ldots, s_n)\, s_i$$

It is also true that

$$\lim_{\substack{t_i \to \infty \\ t_j \to \infty}} y(t_1, \ldots, t_n) = \lim_{\substack{s_i \to 0 \\ s_j \to 0}} Y(s_1, \ldots, s_n)\, s_i s_j$$

and so on, for any number of variables. This is proved in Appendix A.3. The conditions for validity are similar to those for the final-value theorem of linear theory. This theorem will be applied in Section V, but there is one use of it that will be mentioned now.

Consider the system of Fig. 16. For the second-order system, $\underline{L}_2$, discussed in section 3.1, we have

$$L_2(s_1, s_2) = \frac{A^2 B}{(s_1+s_2+\beta)(s_1+\alpha)(s_2+\alpha)}$$

Let the input be $x(t) = y(t) + z(t)$, when $z(t)$ is a unit step that starts at $t = -\infty$, and $y(t)$ is some input that starts at $t = 0$. Then the output $f_2(t)$ is given by

$$f_2 = \underline{L}_2(y^2) + \underline{L}_2(yz) + \underline{L}_2(z^2) \tag{173}$$

Since the system has reached steady state before y(t) is put in, Eq. 173 shows that the system, as far as the input y(t) is concerned, is of the form

$$f_2 = \underline{H}_2(y^2) + \underline{H}_1(y) + \underline{H}_o$$

where $\underline{H}_o$ is a zero-order system – that is, a constant – and $\underline{H}_2 = \underline{L}_2$. Hence

$$H_2(s_1, s_2) = \frac{A^2B}{(s_1+s_2+\beta)(s_1+\alpha)(s_2+\alpha)}$$

Now, $2\underline{L}_2(yz) = \underline{H}_1(y)$, and by applying the limit theorem, we find that

$$H_1(s) = \lim_{p\to 0} 2L_2(s,p)\frac{p}{p}$$

$$= \frac{2A^2B}{\alpha}\,\frac{1}{(s+\beta)}\,\frac{1}{(s+\alpha)}$$

The final-value theorem also gives h_o, the constant associated with the system $\underline{H}_o$, and

$$h_o = \lim_{\substack{s_1\to 0\\ s_2\to 0}} L_2(s_1, s_2) = \frac{A^2B}{\alpha^2\beta}$$

This problem introduces two concepts: (a) the idea of describing a system about a dc input, and (b) the use of this modification of the final-value theorem to find the transforms of the new system. In general, a system of any order can be considered in this fashion.

4.4 DELAY THEOREM

The delay theorem states that if the system $\underline{T}$ is a pure delay (or advance), with $y(t-T) = \underline{T}[y(t)]$, then, for any nonlinear system $\underline{H}$, $\underline{T} * \underline{H} = \underline{H} * \underline{T}$.

This follows from the physical reason that it does not matter if a time delay precedes or follows a system operation of any kind. The particular case

$$\underline{T} * \underline{H}_n = \underline{H}_n * \underline{T}$$

can be derived from transform theory because

$$e^{-s_1T-\dots-s_nT}\, H_n(s_1, \dots, s_n) \tag{174}$$

is the transform of $\underline{T} * \underline{H}_n$, and

$$H_n(s_1, \ldots, s_n)\, e^{-s_1T-\ldots-s_nT} \qquad (175)$$

is the transform of $\underline{H}_n * \underline{T}$. Obviously, Eqs. 174 and 175 are equal.

4.5 DIFFERENTIATION THEOREM

If $\underline{D}$ is a differentiating system with

$$\underline{D}[y(t)] = \frac{d}{dt}\, y(t)$$

then

$$\underline{D} * \underline{H}_n = n\underline{H}_n \circ (\underline{D}\cdot\underline{I}^n) \qquad (176)$$

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

We shall prove this by using transform theory. The transform of $\underline{D}$ is s, and the transform of $\underline{I}$ is 1, and hence from Eqs. 90 and 95, the transform of $n\underline{H}_n \circ (\underline{D}\cdot\underline{I}^n)$ is

$$nH_n(s_1, \ldots, s_n)\, s_1 \qquad (177)$$

Applying the symmetrization procedure of section 2.9 gives Eq. 177 in symmetrical form:

$$H_n(s_1, \ldots, s_n)(s_1+\ldots+s_n)$$

The transform of $\underline{D} * \underline{H}_n$ is

$$(s_1+\ldots+s_n)\, H_n(s_1, \ldots, s_n)$$

by application of Eq. 92. Since

$$H_n(s_1, \ldots, s_n)(s_1+\ldots+s_n) = (s_1+\ldots+s_n)\, H_n(s_1, \ldots, s_n)$$

it follows that Eq. 176 is true.

4.6 LIMIT CYCLES

A feedback system (see Fig. 23) for which the total system operation around the loop is $\underline{L}$, with $\underline{L} = \underline{Q} * \underline{P}$, is in force-free (no driving input) balance, in the steady state, when

$$\underline{L}[x] = x \qquad (178)$$

in the steady state. The particular functions $x(t)$ that satisfy Eq. 178 are called "limit cycles." It is seen that $x(t) = 0$ satisfies Eq. 178 ($\underline{L}$ is assumed to have no zero-order part). Therefore all systems have at least one limit cycle. If Eq. 178 has one or more nonzero solutions, then the system output, under appropriate initial excitation conditions,

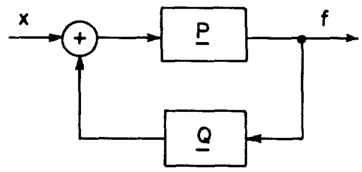


Fig. 23. Feedback system.

will tend toward some one of these solutions in the steady state. A system is, then, unstable if nonzero limit cycles exist.

Some of the cycles themselves may be unstable – that is, the output will tend away from these unstable limit cycles, rather than toward them. If $x(t) = 0$ is an unstable limit cycle, then the system is small-signal unstable. That is, any small signal will cause a system excitation that will not die down.

Returning to the balance equation (Eq. 178), we let $\underline{L} = \underline{H}_1 * \underline{K}$, where $\underline{H}_1$ is a linear lowpass system, and $\underline{K}$ is a nonlinear system. In this case, the balance equation in the steady state can be solved by assuming that

$$x(t) = X \cos \omega t \tag{179}$$

Note that it does not matter if cascade components that make up $\underline{L}$ form a cyclic permutation. For example, $\underline{A} * \underline{B} * \underline{C}$, $\underline{B} * \underline{C} * \underline{A}$, and $\underline{C} * \underline{A} * \underline{B}$ are equivalent forms of $\underline{L}$, as far as Eq. 178 is concerned. All we are doing is writing the balance condition at a different point in the loop. The particular form used is determined by finding out which form gives the easiest answer.

Following the solution of the balance equation, we have

$$\underline{K}[x(t)] = K(X, \omega) \cos \omega t + \text{higher harmonics}$$

where $K(X, \omega)$ is a function of the amplitude X and frequency ω. Because of the lowpass character of $\underline{H}_1$, the solution

$$H_1(j\omega) K(X, \omega) = 1$$

for X and ω is a closely approximate solution of the balance equation. This is the "describing function method" (8), and the value(s) of X and ω, so found, give Eq. 179 as the limit cycle(s). A limit cycle is stable if changing the amplitude X to $X + \Delta X$ gives

$$H_1(j\omega) K(X+\Delta X, \omega) < 1 \qquad \text{for } X + \Delta X > X$$

and

$$H_1(j\omega) K(X+\Delta X, \omega) > 1 \qquad \text{for } X + \Delta X < X$$

Otherwise, the limit cycle is unstable.

For any system in which the loop operation $\underline{L}$ can be described by the functional series, or polynomial, the transforms $H_n(j\omega_1, \ldots, j\omega_n)$ can be used to solve the balance equation (Eq. 178) in the steady state, at least if the number of harmonics involved is not too large. It can be assumed that

$$2x(t) = X_1 e^{j\omega t} + \overline{X}_1 e^{-j\omega t} + X_2 e^{j2\omega t} + \overline{X}_2 e^{j2\omega t} + \ldots$$

and the values of X_1, X_2, etc., and of ω that satisfy Eq. 178 can be found. The x(t) so found are limit cycles and their stability can be investigated, as in the previous special case, by finding the effect of a small amplitude change.

4.7 MEASUREMENT OF NONLINEAR SYSTEMS

Our final topic here is the measurement of nonlinear systems. Two techniques will be mentioned — time-domain and frequency-domain measurements. The discussion of the measurement topic will be completed, in Section V, by describing a measurement procedure based on a white Gaussian-noise input. The discussion here shows only that measurements are theoretically possible. Thus far, no such measurements have been made.

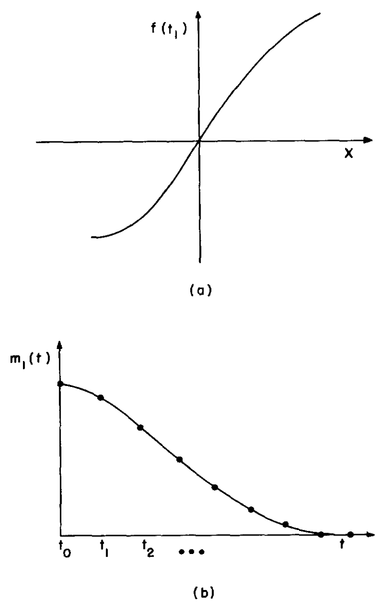


Fig. 24. Measurements: (a) determination of coefficients; (b) linear coefficient $m_1(t)$.

Unlike the input-signal amplitude in linear systems, the amplitude of the input signal of a nonlinear system is of great importance. Both the analysis and measurement of a nonlinear system are dependent on the amplitude range of the input signals for which the system is to be used. For this reason, the input test signals should be bounded signals, and, furthermore, the amplitude of these signals need cover only the range that is of interest.

For the reasons mentioned, we shall adopt the step function for the input test signal for time-domain measurements. Consider a nonlinear system $\underline{L}$, with $\underline{L} = \underline{L}_1 + \underline{L}_2 + \dots + \underline{L}_n + \dots$, and $f = \underline{L}[x]$. The output f(t) for an input step function, x(t) = Xu(t), is

$$f(t) = X \int_0^t l_1(\tau)\, d\tau + \dots + X^n \int_0^t \dots \int_0^t l_n(\tau_1, \dots, \tau_n)\, d\tau_1 \dots d\tau_n + \dots$$

For a particular value of time, t_1,

$$f(t_1) = Xm_1(t_1) + X^2 m_2(t_1) + \dots + X^n m_n(t_1) + \dots \qquad (180)$$

where

$$m_n(t_1) = \int_0^{t_1} \dots \int_0^{t_1} l_n(\tau_1, \dots, \tau_n)\, d\tau_1 \dots d\tau_n$$

Equation 180 is a Taylor series in X, and the output $f(t_1)$ is dependent upon X, as shown in Fig. 24a. If $f(t_1)$ is found experimentally as a function of X, then it is theoretically possible to isolate the coefficients $m_n(t_1)$ in the Taylor's series. If these coefficients are obtained at a set of times $t_1, \ldots, t_n$, then they can be plotted, as is shown in Fig. 24b for $m_1(t)$, to determine $m_n(t)$.

The impulse response of $\underline{L}_1$, can be shown to be the derivative of $m_1(t)$, and so

$$l_1(t) = \frac{d}{dt} m_1(t)$$

Therefore, the impulse response $l_1(t)$ can be theoretically determined.

Now, the impulse response of $\underline{L}_2$ can be found. To do this, we take as input $x(t) = y(t) + z(t)$, where $y(t) = Xu(t)$, $z(t) = Xu(t+T)$, and T is some positive number. The output, then, is

$$f(t) = Xp_1(t) + X^2 p_2(t) + \ldots + X^n p_n(t) + \ldots \tag{181}$$

where

$$p_n(t) = \int_0^t \cdots \int_0^t l_n(\tau_1, \ldots, \tau_n)\, X(t-\tau_1) \ldots x(t-\tau_n)\, d\tau_1 \ldots d\tau_n$$

Again, the $p_n(t)$ can be determined by the use of Taylor's series, as the $m_n(t)$ were. The term $Xp_1(t)$ is not needed and can be ignored. From $p_2(t)$ the impulse response $l_2(t_1, t_2)$ can be found in the following way.

$$\begin{aligned} p_2(t) &= \underline{L}_2((y+z)^2) \\ &= \underline{L}_2(y^2+2zy+z^2) \\ &= \int_0^t \int_0^t l_2(\tau_1, \tau_2)\, d\tau_1 d\tau_2 + 2\int_0^t \int_0^{t+T} l_2(\tau_1, \tau_2)\, d\tau_1 d\tau_2 \\ &\quad + \int_0^{t+T} \int_0^{t+T} l_2(\tau_1, \tau_2)\, d\tau_1 d\tau_2 \\ &= m_2(t) + 2\int_0^t \int_0^{t+T} l_2(\tau_1, \tau_2)\, d\tau_1 d\tau_2 + m_2(t+T) \end{aligned}$$

But, $m_2(t)$ is known, and so the term

$$g_2(t, t+T) = \int_0^t \int_0^{t+T} l_2(\tau_1, \tau_2)\, d\tau_1 d\tau_2 \tag{182}$$

can be isolated. Repeating this measurement for a number of values of T will produce the two-variable function $g_2(t_1, t_2)$. Then it can be shown that

$$l_2(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} g_2(t_1, t_2)$$

and now $l_2(t_1, t_2)$ can be theoretically determined from $g_2(t_1, t_2)$.

In a similar manner, $l_3(t_1, t_2, t_3)$ can be found by using an input triplet of step functions, that is

$$x(t) = Xu(t) + Xu(t+T_1) + Xu(t+T_2)$$

Theoretically, the procedure can be continued to find the $l_n(t_1, \ldots, t_n)$ to any order n that is desired.

The frequency-domain measurements are similar to the time-domain measurements, in their use of Taylor's series to isolate the various terms. We have

$$f(t) = \underline{H}_1[x(t)] + \underline{H}_2[x(t)] + \ldots + \underline{H}_n[x(t)] + \ldots \tag{183}$$

where $\underline{H}_1$ has a transform $H_1(j\omega)$, $\underline{H}_2$ has a transform $H_2(j\omega)$, and so on. Let the input $x(t)$ be a sinusoid, and then

$$x(t) = X \operatorname{Re} \{e^{j\omega t}\}$$

where X is a real number. Direct application of the steady-state methods of section 3.3 gives an output that is the real part of

$$\begin{aligned} f(t) = {} & XH_1(j\omega)\, e^{j\omega t} + \frac{1}{2} X^2 H_2(j\omega, -j\omega) \\ & + \frac{1}{2} X^2 H_2(j\omega, j\omega)\, e^{j2\omega t} \\ & + \frac{3}{4} X^3 H_3(j\omega, j\omega, -j\omega)\, e^{j\omega t} \\ & + \frac{1}{4} X^3 H_3(j\omega, j\omega, j\omega)\, e^{j3\omega t} \\ & + \frac{3}{8} X^4 H_4(j\omega, j\omega, -j\omega, -j\omega) + \ldots \end{aligned} \tag{184}$$

Steady-state harmonic measurements can be taken to determine the coefficients of $e^{jn\omega t}$, which are

$$\frac{1}{2} X^2 H_2(j\omega, -j\omega) + \frac{3}{8} X^4 H_4(j\omega, j\omega, -j\omega, -j\omega) + \ldots$$

for $n = 0$, and

$$XH_1(j\omega) + \frac{3}{4} X^3 H_3(j\omega, j\omega, -j\omega) + \ldots$$

for $n = 1$, and so on. With measurements for various frequencies and values of X, the Taylor-series approach can be used to isolate $H_1(j\omega)$, $H_2(j\omega, j\omega)$, $H_2(j\omega, -j\omega)$, and so forth. In a manner quite similar to the previous use of multiple-step inputs, input sinusoids of the form

$$X e^{j\omega_1 t} + X e^{j\omega_2 t}$$

are used to obtain $H_2(j\omega, j\omega_2)$. In general, multiple sinusoidal inputs can be used to determine the $H_n(j\omega_1, \ldots, j\omega_n)$.

Two methods have been described for the determination of the impulse responses or transforms that characterize a nonlinear system. In Section V, another method, based on a random input, will be discussed.

Note that the measurement of impulse responses and transforms is considerably more complicated than such measurements for linear systems. This is to be expected.

V. RANDOM INPUTS

5.1 INTRODUCTION

In Sections I-IV we have been concerned with the functional representation of continuous nonlinear systems, and with the use of this representation in conjunction with deterministic inputs. We shall now consider random inputs. Output averages and correlation functions will be computed by means of the functional representation. Gaussian inputs will receive the principal emphasis, and certain optimum operations on Gaussian and Gaussian-derived signals will be developed. A system-measurement technique based on a white noise input will be discussed.

5.2 OUTPUT AVERAGES

Let us consider $f = \underline{H}[x]$, where $\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n + \dots$. A typical term is

$$f_n(t) = \underline{H}_n[x(t)]$$

$$= \int \dots \int h_n(\tau_1, \dots, \tau_n)\, x(t-\tau_1) \dots x(t-\tau_n)\, d\tau_1 \dots d\tau_n \tag{185}$$

and

$$f(t) = f_1(t) + \dots + f_n(t) + \dots \tag{186}$$

Now, taking averages on both sides of Eq. 186, we have

$$\overline{f(t)} = \overline{f_1(t)} + \dots + \overline{f_n(t)} + \dots$$

and the object is to find $\overline{f(t)}$ by computing the $\overline{f_n(t)}$. (Here, we consider all random signals to be ergodic. Therefore, averages can be taken as time averages or ensemble averages. The average of a signal z(t) will be denoted $\overline{z(t)}$.) This $\overline{f_n(t)}$ is given by

$$\overline{f_n(t)} = \overline{\int \dots \int h_n(\tau_1, \dots, \tau_n)\, x(t-\tau_1) \dots x(t-\tau_n)\, d\tau_1 \dots d\tau_n} \tag{187}$$

Interchanging orders of integration and averaging in Eq. 187 gives

$$\overline{f_n(t)} = \int \dots \int h_n(\tau_1, \dots, \tau_n)\, \overline{x(t-\tau_1) \dots x(t-\tau_n)}\, d\tau_1 \dots d\tau_n \tag{188}$$

If the correlation function $\overline{x(t_1) \dots x(t_n)}$ is known, $\overline{f_n(t)}$ can be found by performing the integrations of Eq. 188.

It is convenient to introduce a short notation that is related to the operator notation used previously. In this notation Eq. 188 becomes

$$\overline{f_n} = \underline{H}_n\overline{(x_1x_2\dots x_n)}$$

and the average of the output f is

$$\overline{f_n} = \sum_n \underline{H}_n \overline{(x_1 \ldots x_n)}$$

The subscripts 1, 2, ..., n refer to the subscripts of $\tau_1, \tau_2, \ldots, \tau_n$ in Eq. 188.

Similarly, for the calculation of output correlation functions, we have

$$\overline{f(t)\, f(t+T)} = \overline{\{f_1(t)+\ldots+f_n(t)+\ldots\}\{f_1(t+T)+\ldots+f_n(t+T)+\ldots\}}$$

$$= \sum_m \sum_n \overline{f_m(t)\, f_n(t+T)} \tag{189}$$

and

$$\overline{f_m(t)\, f_n(t+T)} = \overline{\int \cdots \int h_m(\tau_1, \ldots, \tau_m)\, x(t-\tau_1) \ldots x(t-\tau_m)}$$

$$\overline{d\tau_1 \ldots d\tau_m \int \cdots \int h_n(\tau_1, \ldots, \tau_n)\, x(t+T-\tau_1)}$$

$$\overline{\ldots x(t+T-\tau_n)\, d\tau_1 \ldots d\tau_n} \tag{190}$$

After rearrangement and interchange of the order of averaging and integrating Eq. 190 becomes

$$\overline{f_m(t)\, f_n(t+T)} = \int \cdots \int h_m(\tau_1, \ldots, \tau_m)\, h_n(T+\tau_{m+1}, \ldots, T+\tau_{m+n})$$

$$\overline{x(t-\tau_1) \ldots x(t-\tau_{m+n})}\, d\tau_1 \ldots d\tau_{m+n} \tag{191}$$

The "impulse response" in this expression is that of the system $\underline{H}_m \cdot (\underline{P}*\underline{H}_n)$, where $\underline{P}$ is an ideal predictor with time shift T, and has an impulse response $\delta(t+T)$. We abbreviate this as $\underline{H}_m \cdot \underline{H}_n^T$, and then Eq. 189, in the short notation, becomes

$$\overline{f_n(t)\, (f_m(t+T)} = \left(\underline{H}_m \cdot \underline{H}_n^T\right) \overline{(x_1 \ldots x_{m+n})} \tag{192}$$

As in the previous case, the output autocorrelation function can be computed if the higher-order input correlation functions are known.

5.3 GAUSSIAN INPUTS

In the important situation in which the input signal is Gaussianly distributed, the calculation of the output averages is not too difficult. Emphasis will be placed on such inputs. First, the special case of white Gaussian inputs will be considered and then this will be generalized. Wiener (5) has rigorously considered the white Gaussian-input case.

If x(t) is white Gaussian with a power density of 1 watt per cycle, then

$$\overline{x(t)} = 0$$

$$\overline{x(t_1)\, x(t_2)} = \delta(t_2-t_1)$$

$$\overline{x(t_1)\, x(t_2)\, x(t_3)} = 0$$

$$\overline{x(t_1)\, x(t_2)\, x(t_3)\, x(t_4)} = \delta(t_2-t_1)\, \delta(t_4-t_3) + \delta(t_3-t_1)\, \delta(t_4-t_2) + \delta(t_4-t_1)\, \delta(t_3-t_2)$$

and so on, where $\delta(t)$ is the unit impulse function. In general, the average is zero if the number of x's is odd, and is a sum of products of impulse responses if the number is even. In general,

$$\overline{x(t_1) \ldots x(t_n)} = \sum \prod_{i,j} \delta(t_i-t_j) \tag{193}$$

The product is over some set of pairs of numbers taken from the numbers 1, 2, ..., n, such as (1, 3), (2, 4), (5, 7), and so forth. The sum is over all such sets.

In the n^{th}-order case, there are $N = (n-1)(n-3) \ldots 1$ terms in the summation, and so for n even

$$\overline{f_n} = \underline{H}_n\overline{(x_1 \ldots x_n)} = \int \cdots \int h_n(\tau_1, \tau_2, \ldots, \tau_n) \sum \prod_{i,j} \delta(\tau_i-\tau_j)\, d\tau_1 \ldots d\tau_n$$

$$= N \int \cdots \int h_n(\tau_1, \tau_1, \tau_2, \tau_2, \ldots, \tau_{n/2}, \tau_{n/2})\, d\tau_1 d\tau_2 \ldots d\tau_{n/2} \tag{194}$$

where $h_n(t_1, \ldots, t_n)$ is symmetrical. (Note that because of this symmetry, the various terms in the sum of Eq. 193 contribute identically in Eq. 194.) Hence, $\overline{f(t)}$ can be determined by performing the integration of Eq. 194 for each of the $\overline{f_n(t)}$ in the sum

$$\overline{f(t)} = \sum_n \overline{f_n(t)} \tag{195}$$

A typical term in the correlation function equation (Eq. 192) is

$$\left(\underline{H}_m \cdot \underline{H}_n^T\right)\overline{(x_1 \ldots x_{m+n})} \tag{196}$$

where $\left(\underline{H}_m \cdot \underline{H}_n^T\right)$ has an impulse response

$$h_m(t_1, \ldots, t_m)\, h_n(T+t_{m+1}, \ldots, T+t_{m+n}) \tag{197}$$

and this impulse response is not symmetrical. Therefore it is necessary to take into account the various terms of Eq. 193. For example,

$$\left(\underline{H}_2 \cdot \underline{H}_2^T\right)\overline{(x_1x_2x_3x_4)} = \iiiint h_2(\tau_1,\tau_2)\,h_2(T+\tau_3,T+\tau_4)$$

$$\{\delta(\tau_2-\tau_1)\delta(\tau_4-\tau_3)+\delta(\tau_3-\tau_1)\delta(\tau_4-\tau_2)+\delta(\tau_4-\tau_1)\delta(\tau_3-\tau_2)\}\, d\tau_1 d\tau_2 d\tau_3 d\tau_4$$

$$= \iint h_2(\tau_1,\tau_1)\,h_2(\tau_2,\tau_2)\, d\tau_1 d\tau_2 + 2\iint h_2(\tau_1,\tau_2)\,h_2(T+\tau_1,T+\tau_2)\, d\tau_1 d\tau_2 \tag{198}$$

In a similar manner, we obtain

$$\left(\underline{H}_1 \cdot \underline{H}_3^T\right)\overline{(x_1x_2x_3x_4)} = \iiiint h_1(\tau_1)\,h_3(T+\tau_2,T+\tau_3,T+\tau_4)$$

$$\{\delta(\tau_2-\tau_1)\delta(\tau_4-\tau_3)+\delta(\tau_3-\tau_1)\delta(\tau_4-\tau_2)+\delta(\tau_4-\tau_1)\delta(\tau_3-\tau_2)\}\, d\tau_1 d\tau_2 d\tau_3 d\tau_4$$

$$= \iint h_1(\tau)\,h_3(T+\tau,\sigma,\sigma)\, d\tau d\sigma \tag{199}$$

Generally, when we are faced with an unsymmetrical situation it is a straightforward matter to determine the various terms of expression 197. The general term that arises is

$$\int \cdots \int h_m(\tau_1,\ldots,\tau_p,\sigma_1,\sigma_1,\ldots,\sigma_q,\sigma_q)$$

$$\times\, h_n(T+\tau_1,\ldots,T+\tau_p,\sigma_{q+1},\sigma_{q+1},\ldots,\sigma_r,\sigma_r)\, d\tau_1 \ldots d\tau_p\, d\sigma_1 \ldots d\sigma_r \tag{200}$$

Here, $p + 2q = m$, and $p + 2r - 2q = n$. It should be remembered that expression 197 equals zero if $m + n$ is odd. Once the terms $\left(\underline{H}_m \cdot \underline{H}_n^T\right)\overline{(x_1 \ldots x_{m+n})}$ have been determined, $\overline{f(t)\ f(t+T)}$ is given by Eq. 192.

The results for white Gaussian inputs can be used to obtain output averages and correlation functions for non-white Gaussian inputs into a system $\underline{H}$. In the non-white case, the Gaussian signal can be formed from a white Gaussian signal by means of a linear shaping filter, $\underline{K}_1$. This is illustrated in Fig. 25. Then, rather than work with a non-white Gaussian input to a system $\underline{H}$, we work with a white Gaussian input to a system $\underline{H} * \underline{K}_1$. Also, if the input to a system $\underline{H}$ is non-Gaussian, but formed from a white Gaussian signal by a known nonlinear operation $\underline{K}$ (which can be expanded in the functional representation), then we can work with a white Gaussian signal to a system $\underline{H} * \underline{K}$.

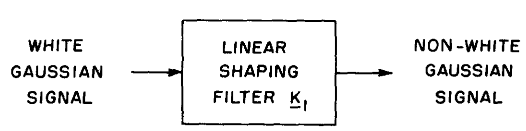


Fig. 25. Illustrating the use of shaping filter.

5.4 USE OF TRANSFORMS

The averages given, for example, by Eqs. 194 and 198 could be found by performing the indicated integrations of the usually awkward impulse responses. However, this difficulty can be overcome by the use of transforms. The transforms considered here will always be Fourier transforms (see Appendix A.1) and $s = j\omega$.

To develop the use of transforms, three typical situations will be explained. First, consider the term

$$\underline{H}_n \overline{(x_1 \ldots x_n)} = \int \cdots \int h_n(\tau_1, \tau_1, \ldots, \tau_{n/2}, \tau_{n/2})\, d\tau_1 \ldots d\tau_{n/2} \tag{201}$$

from Eq. 194. The transform of $h_n(t_1, \ldots, t_n)$ is $H_n(s_1, \ldots, s_n)$, and hence the transform of $h_n(t_1, t_1, t_3, \ldots, t_n)$ can be obtained by inspection if $H_n(s_1, \ldots, s_n)$ is factorizable.

Let the transform of $h_n(t_1, t_1, t_3, \ldots, t_n)$ be

$$K_{n-1}(s_1, s_3, \ldots, s_n)$$

Now, the first integration of Eq. 201 can be performed. This integration is

$$\int h_n(\tau_1, \tau_1, \tau_3, \ldots, \tau_n)\, d\tau_1$$

and it can be obtained from $K_{n-1}(s_1, \ldots, s_n)$ by the method of section 4.4. That is, the transform is

$$\lim_{s_1 \to 0} K_{n-1}(s_1, s_3, \ldots, s_n)$$

Let this expression equal $L_{n-2}(s_3, \ldots, s_n)$, which is the transform of

$$\int h_n(\tau_1, \tau_1, \tau_3, \ldots, \tau_n)\, d\tau_1$$

The operation can be repeated on $L_{n-2}(s_3, \ldots, s_n)$ to perform the second integration of Eq. 201, and so on until it has been evaluated.

As an example, consider

$$\iint l_4(\tau_1, \tau_1, \tau_2, \tau_2)\, d\tau_1 d\tau_2 \tag{202}$$

where $l_4(t_1, t_2, t_3, t_4)$ has a transform

$$\frac{A}{s_1 + s_2 + s_3 + s_4 + \alpha}\, \frac{1}{s_1 + \beta}\, \frac{1}{s_2 + \beta}\, \frac{1}{s_3 + \beta}\, \frac{1}{s_4 + \beta} \tag{203}$$

The first association gives

$$\frac{A}{s_1 + s_3 + s_4 + \alpha} \frac{1}{s_1 + 2\beta} \frac{1}{s_3 + \beta} \frac{1}{s_4 + \beta}$$

and taking the limit $s_1 \to 0$ gives

$$\frac{A}{s_3 + s_4 + \alpha} \frac{1}{2\beta} \frac{1}{s_3 + \beta} \frac{1}{s_4 + \beta}$$

Associating the other two variables, we obtain

$$\frac{1}{2\beta} \frac{A}{s + \alpha} \frac{1}{s + 2\beta} \tag{204}$$

and

$$\iint l_4(\tau_1, \tau_1, \tau_2, \tau_2)\, d\tau_1 d\tau_2 = \lim_{s\to 0} \frac{1}{2\beta} \frac{A}{s + \alpha} \frac{1}{s + 2\beta} = \frac{A}{4\alpha\beta^2} \tag{205}$$

The second situation to be studied is

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_n)\, k_n(T+\tau_1, \ldots, T+\tau_n)\, d\tau_1 \ldots d\tau_n \tag{206}$$

As we have done before with transforms, we introduce $T_1, \ldots, T_n$ into this term and consider

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_n)\, k_n(T_1+\tau_1, \ldots, T_n+\tau_n)\, d\tau_1 \ldots d\tau_n$$

Taking the higher-order transform of this expression yields

$$H_n(-s_1, \ldots, -s_n)\, K_n(s_1, \ldots, s_n) \tag{207}$$

The actual transform of Eq. 206 can now be obtained from Eq. 207 by associating $T_1, \ldots, T_n$ with T by means of the inspection technique if the transforms are factorizable.

In using the inspection technique it should be noted that the contribution of terms of the form

$$L_{n+1}(s_1, \ldots, s_n, s_i+s_j)\, P(s_i)\, Q(-s_j) \tag{208}$$

is zero when T_i and T_j are associated. This is so because the T_i and T_j in the inverse transform of Eq. 208 are in disjoint regions; that is, $T_i > 0$ and $T_j < 0$. Hence there is no contribution for $t = T_i = T_j$.

In order to illustrate the method, consider the case in which

$$H_2(s_1,s_2)\, H_2(-s_1,-s_2) = \frac{K^2}{(-s_1-s_2+\beta)(s_1+s_2+\beta)(-s_1+\alpha)(-s_2+\alpha)}\, \frac{1}{(s_1+\alpha)(s_2+\alpha)}$$

$$= \frac{1}{4\alpha^2}\, \frac{K^2}{(-s_1-s_2+\beta)(s_1+s_2+\beta)} \left\{\frac{1}{-s_1+\alpha} + \frac{1}{s_1+\alpha}\right\} \left\{\frac{1}{-s_2+\alpha} + \frac{1}{s_2+\alpha}\right\}$$

$$= \frac{1}{4\alpha^2}\, \frac{K^2}{(-s_1-s_2+\beta)(s_1+s_2+\beta)} \left\{\frac{1}{(-s_1+\alpha)(-s_2+\alpha)} + \frac{1}{(s_1+\alpha)(s_2+\alpha)} + \frac{1}{(s_1+\alpha)(-s_2+\alpha)} + \frac{1}{(-s_1+\alpha)(s_2+\alpha)}\right\} \tag{209}$$

Associating the variables by inspection yields

$$\frac{K^2}{4\alpha^2}\, \frac{1}{(-s+\beta)(s+\beta)} \left\{\frac{1}{-s+2\alpha} + \frac{1}{s+2\alpha}\right\} \tag{210}$$

and the terms involving $[(s_1+\alpha)(-s_2+\alpha)]^{-1}$, and $[(-s_1+\alpha)(s_2+\alpha)]^{-1}$ give no contribution. Equation 210 is then the transform of $\iint h_2(\tau_1, \tau_2)\, h_2(T+\tau_1, T+\tau_2)\, d\tau_1 d\tau_2$, where $h_2(t_1, t_2)$ has the transform $H_2(s_1, s_2)$.

A third situation that arises is

$$\int \cdots \int h_n(\tau_1, \ldots, \tau_n)\, k_m(\tau_1+T, \ldots, \tau_n+T, \sigma_1, \sigma_1, \ldots, \sigma_p, \sigma_p)\, d\tau_1 \ldots d\tau_n\, d\sigma_1 \ldots d\sigma_p$$

where m + n is even, p = (m−n)/2, and m > n. First, consider

$$\int \cdots \int k_m(t_1, \ldots, t_n, \sigma_1, \sigma_1, \ldots, \sigma_p, \sigma_p)\, d\sigma_1 \ldots d\sigma_p \tag{211}$$

where $k_m(t_1, \ldots, t_m)$ has a transform $K_m(s_1, \ldots, s_m)$. By direct application of the first method discussed in this section, the transform of Eq. 211 can be obtained. Once this has been done, the situation is the same as in the second case and the method involved there can be used. In a similar manner, the general form of Eq. 200 can be handled.

For example, consider

$$\iint h_1(\tau)\, h_3(T+\tau, \sigma, \sigma)\, d\tau d\sigma \tag{212}$$

where $h_3(t_1, t_2, t_3)$ has a transform

$$\frac{K}{s_1 + s_2 + s_3 + \alpha}\, \frac{1}{s_1 + \beta}\, \frac{1}{s_2 + \beta}\, \frac{1}{s_3 + \beta} \tag{213}$$

and $h_1(t)$ has a transform $H/(s+\alpha)$.

Consider

$$\int h_3(t, \sigma, \sigma)\, d\sigma \tag{214}$$

Associating s_2 and s_3 in Eq. 213 yields

$$\frac{K}{s_1 + s_2 + \alpha} \frac{1}{s_1 + \beta} \frac{1}{s_2 + 2\beta} \tag{215}$$

Let $s_2 \to 0$, in order to evaluate the integral of Eq. 214, and we have

$$\frac{K}{s_1 + \alpha} \frac{1}{s_1 + \beta} \frac{1}{2\beta} \tag{216}$$

which is the transform of Eq. 214. Equation 212 has a transform

$$\frac{KH}{2\beta} \frac{1}{(-s+\alpha)} \frac{1}{(s+\alpha)(s+\beta)}$$

Inverting this transform gives

$$\frac{1}{2(\alpha+\beta)(\alpha-\beta)} \{2\alpha e^{-\beta t} - (\alpha+\beta) e^{-\alpha t}\} \qquad \text{for } t > 0$$

and

$$\frac{1}{2\alpha(\alpha+\beta)} e^{\alpha t}, \qquad \text{for } t < 0$$

for Eq. 212.

The three main methods for handling the expressions that arise in computing output averages or correlation functions have been presented. These are transform or frequency-domain methods. In computing autocorrelation functions the results can be left in the frequency domain, in which they represent the spectra.

5.5 EXAMPLE 5.

The system for this example is shown in Fig. 26. It is an apparatus for measuring the average square of the Gaussian signal, y(t). The signal y(t) is formed from a white Gaussian x(t) by means of the shaping filter $\underline{A}_1$, with $A_1(s) = A/(s+\alpha)$. The system $\underline{B}_1$ is a physical approximation to an ideal integrator, and $B_1(s) = B/(s+\beta)$. The over-all system operating on x(t) is $\underline{L}_2 = \underline{B}_1 * \underline{N}_2 * \underline{A}_1$, where $\underline{N}_2$ is a no-memory squarer, and $n_2 = 1$.

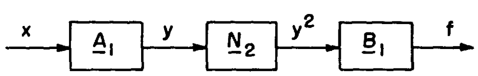


Fig. 26. Apparatus for measuring average square of the signal y(t).

First, we shall obtain the average square of y(t). This is the average of the output of the system $\underline{H}_2 = \underline{N}_2 * \underline{A}_1$

operating on x(t). The output from $\underline{\underline{H}}_2$ is $y^2(t)$. Then

$$\overline{y^2(t)} = \underline{\underline{H}}_2\overline{(x_1x_2)}$$

$$= \iint h_2(\tau_1, \tau_2)\, \delta(\tau_2-\tau_1)\, d\tau_1 d\tau_2$$

$$= \int h_2(\tau, \tau)\, d\tau$$

and $h_2(t_1, t_2)$ has a transform,

$$H_2(s_1, s_2) = \frac{A^2}{(s_1+a)(s_2+a)}$$

Associating t_1 and t_2 gives $A^2/(s+2a)$, from Eq. 203, and so

$$\overline{y^2(t)} = \lim_{s\to 0} \frac{A^2}{s+2a} = \frac{A^2}{2a}$$

Next, we shall obtain the average output from the system $\underline{\underline{L}}_2$, where

$$L_2(s_1, s_2) = \frac{A^2B}{(s_1+s_2+\beta)(s_1+a)(s_2+a)} \tag{217}$$

Associating the variables gives

$$\frac{A^2B}{(s+\beta)(s+2a)}$$

from Eq. 204. Then

$$\overline{f(t)} = \lim_{s\to 0} \frac{A^2B}{(s+\beta)(s+2a)} = \frac{A^2}{2a}\frac{B}{\beta}$$

and $\overline{f(t)} = \overline{y^2(t)}$ when $B = \beta$. We see that the apparatus does measure the average of $y^2(t)$. However, the output f(t) is not a constant, but a random variable.

To conclude this example, we obtain the output spectrum $\Phi_f(\omega)$. The output autocorrelation is given by application of Eq. 199:

$$\Phi_f(T) = \iint l_2(\tau_1, \tau_1)\, l_2(\tau_2, \tau_2)\, d\tau_1 d\tau_2$$

$$+ 2\iint l_2(\tau_1, \tau_2)\, l_2(T+\tau_1, T+\tau_2)\, d\tau_1 d\tau_2 \tag{218}$$

where $l_2(t_1, t_2)$ is the impulse response of the system. The first term of Eq. 218 can be shown (see Appendix B.3) to equal $(A^2/2a)^2$, when $B = \beta$. Also, from Eq. 217,

$$2L_2(-s_1,-s_2)\, L_2(s_1,s_2) = \frac{2A^4B^2}{(-s_1-s_2+\beta)(s_1+s_2+\beta)(-s_1+a)(-s_2+a)(s_1+a)(s_2+a)} \tag{219}$$

and application of the inspection method (see Appendix B.3) gives

$$\frac{2A^4B^2}{\omega^2+\beta^2}\,\frac{1}{\omega^2+4a^2}$$

when $j\omega = s$. Therefore, the spectrum $\Phi_f(\omega)$ is given by

$$\Phi_f(\omega) = \left(\frac{A^2B}{2a\beta}\right)^2 \delta(\omega) + \frac{2A^4B^2}{\omega^2+\beta^2}\,\frac{1}{\omega^2+4a^2} \tag{220}$$

5.6 EXAMPLE 6.

This example is concerned with the feedback system of section 3.5. The problem is much the same, except that here the input is a random signal, Gaussianly distributed. Our object is to use the nonlinear compensating device $\underline{N}$ to decrease the servo following error and still meet a constraint on the maximum allowable rms acceleration.

The system input is x(t), the output is f(t), and the following error is $e(t) = x(t) - f(t)$. The acceleration of the motor is $a(t) = d/dt\, f(t)$.

First, we shall consider the linear, uncompensated system with $\underline{N} = \underline{I}$. The input spectrum is

$$\Phi_x(\omega) = \frac{B^2}{\omega^2+\beta^2} \tag{221}$$

The pertinent results (see Appendix B.4) are:

$$\overline{e^2} = \frac{B^2}{2A} \tag{222}$$

$$\overline{a^2} = \frac{AB^2}{2} \tag{223}$$

$$\overline{f^2} = \frac{B^2}{\beta} \tag{224}$$

$$\frac{\overline{e^2}}{\overline{f^2}} = \frac{\beta}{2A} \tag{225}$$

where $A > 10\beta$, in order that the following error be small. If M is the maximum allowable rms acceleration, then minimum following error is obtained for $A = 2M^2/B^2$.

The results for the compensated nonlinear system with $\underline{N} = \underline{I} + \underline{N}_3$ (see Appendix B.4) will also be stated:

$$\overline{e^2} = \frac{B^2}{2A} + \frac{3}{4} n_3 \frac{B^4}{A^2} + \frac{3}{4} n_3^2 \frac{B^6}{A^3}$$

$$\overline{a^2} = \frac{AB^2}{2} + \frac{3}{4} n_3 \frac{B^4}{A} + \frac{21}{32} n_3^2 \frac{B^6}{A^2}$$

and

$$\overline{f^2} = \frac{B^2}{\beta} + \frac{3}{4} n_3 \frac{B^4}{A^2} + \frac{3}{4} n_3^2 \frac{B^6}{A^4}$$

where n_3 is taken sufficiently small that only the first three terms are significant. Now, for example, we shall take the numerical values: $B^2 = 1/2$, $\beta = 2/30$, and $M^2 = 5/32$. Then for the linear case, the largest allowable $A = 20/32$ and $\overline{e^2}/\overline{f^2} = 0.05$.

For the nonlinear compensated case, with $n_3 = -\frac{1}{2}$, the allowable A is 1, and $\overline{e^2}/\overline{f^2} = 0.021$. This represents a 60 per cent decrease in the following error.

This example shows what can be done by applying the functional representation to nonlinear systems with random Gaussian inputs. It also illustrates the possible use of nonlinear elements for servo compensation when the input is a random signal.

5.7 OPTIMUM SYSTEMS

This section deals with the problem of obtaining the realizable system that best approximates a desired unrealizable nonlinear system or operation. The desired system is unrealizable because its impulse response "starts before t = 0." Best is to be taken in the least-mean-square sense; that is, the average squared error between the output of the realizable system and the output of the unrealizable system is minimum. The signals upon which the systems operate are Gaussian. Barrett (6) has developed an approach for the general (non-Gaussian) signal, but there are problems still to be solved before we can take advantage of his approach. In this report we are restricted to Gaussian signals or signals derived from Gaussian signals.

We shall consider an unrealizable linear system $\underline{H}_1$ with a white Gaussian input, and we shall find the optimum realizable system. Let the impulse response $h_1(t)$ be nonzero for $t < 0$. Then

$$f_1(t) = \int_{-\infty}^{\infty} h_1(\tau)\, x(t-\tau)\, d\tau$$

or, if we divide the region of integration into two parts, we have

$$f_1(t) = \int_{-\infty}^{0} h_1(\tau)\, x(t-\tau)\, d\tau + \int_{0}^{\infty} h_1(\tau)\, x(t-\tau)\, d\tau$$

The region $-\infty$ to 0 covers x(t) in the future, and the region 0 to ∞ covers x(t) in the past. If we know only the past of x(t), then $f_1(t)$ can only be estimated. The best mean-square estimate of f(t) is $g_1(t)$, with

$$g_1(t) = \text{average of } f_1(t) \text{ over the future} \tag{226}$$

or

$$g_1(t) = \int_{-\infty}^{0} h_1(\tau)\, \overline{x(t-\tau)}\, d\tau + \int_{0}^{\infty} h_1(\tau)\, x(t-\tau)\, d\tau \tag{227}$$

Equation 227 follows from Eq. 226 because x(t) is white, and therefore the past and future of x(t) are uncorrelated. Since $\overline{x(t)} = 0$ for a Gaussian (zero-mean) signal, Eq. 226 becomes

$$g(t) = \int_{0}^{\infty} h_1(\tau)\, x(t-\tau)\, d\tau \tag{228}$$

and the best estimate, g(t), has been found. Putting this another way: the unrealizable impulse response $h_1(t)$ has been replaced by $k_1(t)$, where

$$k_1(t) = \begin{cases} h_1(t) & \text{for } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_1(t) = \int_{-\infty}^{\infty} k_1(\tau)\, x(t-\tau)\, d\tau \tag{229}$$

This is a familiar result from linear theory.

Now we shall do the same thing for a second-order system $\underline{H}_2$, in which

$$f_2(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2$$

Splitting the region of integration into past and future regions gives

$$\begin{aligned} f_2(t) = &\int_{-\infty}^{0} \int_{-\infty}^{0} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \\ &+ 2\int_{-\infty}^{0} \int_{0}^{\infty} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \\ &+ \int_{0}^{\infty} \int_{0}^{\infty} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2 \end{aligned}$$

The factor 2 in the second term is obtained by taking advantage of the symmetry of $h_2(t_1, t_2)$ and combining two terms. Again, $g_2(t)$, the best estimate of $f_2(t)$, is obtained by averaging over the future. Since x(t) is white Gaussian, the past and future are independent, and

$$g_2(t) = \int_{-\infty}^{0} \int_{-\infty}^{0} h_2(\tau_1, \tau_2) \overline{x(t-\tau_1)\, x(t-\tau_2)}\, d\tau_1 d\tau_2$$

$$+ 2\int_{-\infty}^{0} \int_{0}^{\infty} h_2(\tau_1, \tau_2) \overline{x(t-\tau_1)\, x(t-\tau_2)}\, d\tau_1 d\tau_2$$

$$+ \int_{0}^{\infty} \int_{0}^{\infty} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2$$

Performing the indicated averages gives

$$g_2(t) = \int_{-\infty}^{0} h_2(\tau, \tau)\, d\tau + \int_{0}^{\infty} \int_{0}^{\infty} h_2(\tau_1, \tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, d\tau_1 d\tau_2$$

Thus the unrealizable system $\underline{H}_2$ has been approximated by a realizable system K, with $\underline{K} = \underline{K}_2 + \underline{K}_o$, and $\underline{K}_2$ has an impulse response

$$k_2(t_1, t_2) = \begin{cases} h_2(t_1, t_2) & \text{for } t_1 \text{ and } t_2 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and $\underline{K}_o$ is a zero-order system (a constant) of value

$$k_o = \int_{-\infty}^{0} h_2(\tau, \tau)\, d\tau$$

In general, this procedure can be used to show that an unrealizable system $\underline{H}_n$ is replaced by a best realizable system $\underline{K}_{(n)}$, with

$$\underline{K}_n = \sum_{r=0}^{n} \binom{n}{r} \underline{K}_{n-r} \tag{230}$$

The $\underline{K}_{n-r}$ have impulse responses $k_{n-r}(t_1, \ldots, t_{n-r})$, and for r odd and all t_i (where $i=1, \ldots, n$),

$$k_{n-r}(t_1, \ldots, t_{n-r}) = 0$$

for r even and all $t_i \geq 0$,

$$= (r-1)(r-3) \ldots 1 \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} h_n(\tau_1, \tau_1, \ldots,$$

$$\tau_{r/2}, \tau_{r/2}, t_1, t_2, \ldots, t_{n-r})\, d\tau_1 \ldots d\tau_{r/2}$$

and for r even and some $t_i < 0$,

$$= 0 \tag{231}$$

Now that the realizable system which is nearest, in the mean-square sense, to a

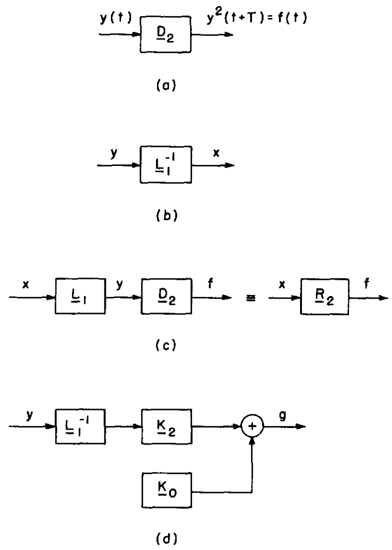


Fig. 27. Example 7. (a) Desired operation. (b) Production of x. (c) Desired operation on x. (d) Optimum system operation on y.

desired unrealizable system has been determined for a white Gaussian input signal, the extension to non-white Gaussian signals can be explained. By means of linear shaping filter $\underline{L}_1$, a non-white Gaussian signal can be whitened. Once this is done, the optimization can proceed with the resultant white Gaussian signal as input. That is, given a desired operation, $\underline{H} = \underline{H}_1 + \dots + \underline{H}_n$, the signal x(t) is whitened by $\underline{L}_1$ to produce y(t), and the optimization is made for a desired system $\underline{H} * \underline{L}_1^{-1}$ with input y(t). The resultant realizable system is $\underline{K}$, and then the optimum system is $\underline{K} * \underline{L}_1$.

A further generalization can be made. Suppose that the input signal is the result of (or the statistical equivalent of) a known nonlinear operation, $\underline{L}$, on a white Gaussian signal. Furthermore, assume that this system $\underline{L}$ has a stable, realizable inverse, $\underline{L}^{-1}$. Then, just as before with $\underline{L}_1$, the optimization procedure can be preceded by the $\underline{L}^{-1}$ operation.

5.8 EXAMPLE 7.

In this example we desire to obtain the best mean-square estimate of $f(t) = y^2(t+T)$, where y(t) is a Gaussian signal, and T is positive. In other words, we want to find the realizable system closest to $\underline{D}_2$, with

$$d_2(t_1, t_2) = \delta(t_1+T)\ \delta(t_2+T) \tag{232}$$

Now, take

$$\Phi_y(\omega) = \frac{D^2(\omega^2+\gamma^2)}{(\omega^2+\alpha^2)(\omega^2+\beta^2)}$$

The shaping filter $\underline{L}_1$ has a transform

$$L_1(s) = \frac{D(s+\gamma)}{(s+\alpha)(s+\beta)}$$

We operate on y(t) with $\underline{L}_1^{-1}$ to produce a white Gaussian x(t), as shown in Fig. 27b. The desired operation on x(t) (see Fig. 27c) is $\underline{R}_2 = \underline{D}_2 * \underline{L}_1$, and

$$R_2(s_1, s_2) = e^{s_1 T} L_1(s_1)\, e^{s_2 T} L_1(s_2)$$

Applying Eqs. 230 and 231 to determine the best realizable operation on a white Gaussian signal gives $\underline{K} = \underline{K}_2 + \underline{K}_o$, and

$$\begin{aligned} k_2(t_1, t_2) &= \left\{ P\, e^{-\alpha(t_1+T)} + Q\, e^{-\beta(t_1+T)} \right\} \\ &= \left\{ P\, e^{-\alpha(t_2+T)} + Q\, e^{-\beta(t_2+T)} \right\} \qquad \text{for } t_1, t_2 \geq 0 \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

and

$$k_o = \left\{ \int_0^T (P\, e^{-\alpha\sigma} + Q\, e^{-\beta\sigma}) d\sigma \right\}^2$$

where $P = D(\gamma-\alpha)/(\beta-\alpha)$, and $Q = D(\gamma-\beta)/(\alpha-\beta)$. The optimum operation on y(t) is, then, $\underline{K} * \underline{L}_1^{-1}$, and the optimum system is shown in Fig. 27d.

5.9 EXAMPLE 8.

This example deals with the prediction of y(t), where $y = (\underline{N} * \underline{L}_1)[x]$; x(t) is a white Gaussian signal; $\underline{N}[z] = z + z^3$; and

$$L_1(s) = \frac{1}{(s+\alpha)(s+\beta)} \tag{233}$$

Operating on y(t) to produce

$$x = \left(\underline{L}_1^{-1} * N^{-1} \right) [y]$$

and performing the optimization on x(t) gives an optimum predictor $\underline{J}$, which operates on y(t), with

$$\underline{J} = \underline{M} * \underline{K}_1 * \underline{N}^{-1} \tag{234}$$

and

$$\underline{M}[z] = m_1 y + y^3 \tag{235}$$

where

$$m_1 = 1 + \frac{3}{(\beta-\alpha)}\left\{\frac{1 - e^{-2\alpha A}}{2\alpha} + \frac{1 - e^{-2\beta A}}{2\beta} + \frac{1 - e^{-(\alpha+\beta)A}}{\alpha + \beta}\right\} \tag{236}$$

$$K_1(s) = c_1 + c_2 s \tag{237}$$

$$c_1 = \frac{\beta e^{-\alpha A} + \alpha e^{-\beta A}}{(\beta-\alpha)} \tag{238}$$

$$c_2 = \frac{e^{-\alpha A} - e^{-\beta A}}{(\beta-\alpha)} \tag{239}$$

See Appendix B.5 for further details.

5.10 THEORETICAL DISCUSSION ON MEASUREMENTS

The way in which the system impulse responses may be obtained by measurements made with a white Gaussian noise input will be demonstrated here. The quantities to be measured will be input-output crosscorrelations. However, at this time, such measurements can only be discussed theoretically.

Let the input x(t) to a linear system $\underline{H}_1$ be white Gaussian noise, and the output be $f_1(t)$. Then

$$\begin{aligned}\overline{f_1(t)\, x(t-T)} &= \int h_1(\tau)\, \overline{x(t-\tau)\, x(t-T)}\, d\tau \\ &= \int h_1(\tau)\, \delta(T-\tau)\, d\tau \\ &= h_1(T)\end{aligned}$$

This method, which is known from linear theory (8), is one means of measuring the impulse response of a linear system.

Now consider a quadratic system $\underline{H}_2$ with input x(t) and output $f_2(t)$. Then

$$\begin{aligned}\overline{f_2(t)\, x(t-T_1)\, x(t-T_2)} &= \iint h_2(\tau_1, \tau_2)\, \overline{x(t-\tau_1)\, x(t-\tau_2)\, x(t-T_1)\, x(t-T_2)}\, d\tau_1 d\tau_2 \\ &= \iint h_2(\tau_1, \tau_2)\{\delta(\tau_2-\tau_1)\delta(T_2-T_1)+\delta(T_1-\tau_1)\delta(T_2-\tau_2) \\ &\quad +\delta(T_2-\tau_1)\delta(T_1-\tau_2)\}\, d\tau_1 d\tau_2 \\ &= \delta(T_2-T_1)\int h_2(\tau, \tau)\, d\tau + h_2(T_1, T_2) + h_2(T_2, T_1)\end{aligned}$$

We note that $\frac{1}{2}\{h_2(T_1,T_2)+H_2(T_2,T_1)\}$ is the symmetrical form for the quadratic impulse response, and so

$$\overline{f_2(t)\, x(t-T_1)\, x(t-T_2)} = \delta(T_2-T_1) \int h_2(\tau,\tau)\, d\tau + 2h_2(T_1,T_2)$$

Measurement of this second-order crosscorrelation, $\overline{f_2(t)\, x(t-T_1)\, x(t-T_2)}$ for $T_1 \neq T_2$, therefore yields $h_2(T_1, T_2)$ for $T_1 \neq T_2$. The function $h_2(T, T)$ can be obtained by crosscorrelating $f_2(t)$ with $\{x^2(t-T)-\overline{x^2(t)}\}$. Then we have

$$\overline{f_2(t)\{x^2(t-T)-\overline{x^2(t)}\}} = 2h_2(T, T) \tag{240}$$

For white noise $\overline{x^2(t)}$ does not exist, but it does exist for any practical approximation to white noise.

For the cubic system $\underline{H}_3$, we have

$$\overline{f_3(t)\, x(t-T_1)\, x(t-T_2)\, x(t-T_3)} = 6h_3(T_1,T_2,T_3) + 3\delta(T_1-T_2) \int h_3(\tau,\tau,T_3)\, d\tau$$

$$+ 3\delta(T_1-T_3) \int h_3(\tau,\tau,T_2)\, d\tau + 3\delta(T_2-T_3) \int h_3(\tau,\tau,T_1)\, d\tau$$

which gives $h_3(T_1, T_2, T_3)$ for $T_1 \neq T_2 \neq T_3$. To obtain $h_3(T_1, T_2, T_3)$ in the excluded region, measurements similar to the measurement indicated in Eq. 240 can be made.

Higher-order systems $\underline{H}_n$ can be handled in an analogous manner. If the measurements are to be made on a system $\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n + \dots$, then to extract the $\underline{H}_1$ term, for example, the measurement of $\overline{f(t)\, x(t-T)}$, where $f(t)$ is the output of $\underline{H}$ with input $Ax(t)$, may be made for an input $Ax(t)$ for different values of the constant, A. Then the part of $\overline{f_1(t)\, x(t-T)}$ that varies as A will be $h_1(T)$.

Similarly, for $\overline{f(t)\, x(t-T_1)\, x(t-T_2)}$, the part that varies as A^2 is $2h_2(T_1, T_2)$, for $T_1 \neq T_2$. Thus, the Taylor's-series method has been applied again to separate the various $\underline{H}_n$ of the system.

5.11 SUMMARY

We have shown how output averages and correlation functions may be obtained for nonlinear systems described by the functional representation. Emphasis has been placed on Gaussian input signals, and frequency-domain techniques have been developed and illustrated by examples.

A discussion was devoted to the problem of optimum nonlinear operations on Gaussian, or Gaussian-derived, signals. Two examples of the optimization procedure were given.

The section on theoretical measurement was intended to show briefly how input-output crosscorrelation measurements can be used to measure the system impulse responses.

VI. THEORETICAL DISCUSSION OF FUNCTIONAL REPRESENTATION

A number of theoretical topics concerning functional representation and the algebra of systems will now be presented. In the first few sections we shall attempt to place certain aspects of this algebra on firmer ground. Then we shall discuss some topics that prescribe theoretical limitations on the functional representation.

6.1 DOMAIN AND RANGE

We have stressed the point that the amplitudes of the input signals of a nonlinear system are very important. They are important for two reasons: (a) the system may act in a radically different way for two signals of the same wave shape but of different amplitudes, and (b) the method of analysis may depend on some limitations on input-signal amplitude. Therefore, to be rigorous, we should associate a certain input limitation with any nonlinear system that is being discussed. This limitation will be the "domain" of the system. For the system $\underline{H}$ it will be denoted D_H, and it is the class of all allowable input signals. If a signal x falls in this class we write $x \in D_H$ (in words, x is contained in D_H). A convenient way to particularly define D_H is to say that there exists a positive number of X that is such that if $|x| < X$, then $x \in D_H$. In general, there are many ways to define the system domain.

If a D_H is defined for a system $\underline{H}$, then the outputs f that are associated with the inputs x, where $x \in D_H$, form a class of signals. This class will be called the "range" of the system, and will be denoted R_H. If $f = \underline{H}[x]$, then $f \in R_H$, for all $x \in D_H$.

A question now arises about what happens when we additively combine two systems of different domains. If $\underline{L} = \underline{J} + \underline{K}$, where $\underline{J}$ has domain D_J and $\underline{K}$ has domain D_K, then we shall take the domain of $\underline{L}$, D_L, to be the class of signals that are contained in both D_J and D_K. Therefore, we shall consider only inputs for $\underline{L}$ that we know are allowable inputs for both $\underline{J}$ and $\underline{K}$.

Similarly, for the multiplication combination $\underline{L} = \underline{J} \cdot \underline{K}$, D_L is the class of signals contained in both D_J and D_K.

For the cascade combination $\underline{L} = \underline{J} * \underline{K}$, we must assume that the range of $\underline{K}$, R_K, is contained in D_J. If this is not so, then D_K must be constrained so that R_K is contained in D_J.

In this report we have assumed that these points were implied when we have combined systems.

6.2 ALGEBRAIC LAWS

The validity of certain operations, which has previously been assumed (see sec. 2.8), will be established here. These operations will be presented as a set of theorems.

THEOREM 1. If $\underline{A} = \underline{B}$ (241)

then $\underline{A} + \underline{H} = \underline{B} + \underline{H}$ (242)

and $\underline{H} + \underline{A} = \underline{H} + \underline{B}$ (243)

This theorem holds also for the minus operation.

THEOREM 2. If $\underline{A} = \underline{B}$

then $\underline{A} \cdot \underline{H} = \underline{B} \cdot \underline{H}$ (244)

and $\underline{H} \cdot \underline{A} = \underline{H} \cdot \underline{B}$ (245)

THEOREM 3. If $\underline{A} = \underline{B}$

then $\underline{H} * \underline{A} = \underline{H} * \underline{B}$ (246)

and $\underline{A} * \underline{H} = \underline{B} * \underline{H}$ (247)

Systems $\underline{A}$, $\underline{B}$, and $\underline{H}$ are nonlinear. These theorems are easily proved.

Let $p = \underline{A}[x]$; $q = \underline{B}[x]$; and $r = \underline{H}[x]$. Then from Eq. 241, $p = q$. But $p + r = q + r$, and so $\underline{A}[x] + \underline{H}[x] = \underline{B}[x] + \underline{H}[x]$ or $\underline{A} + \underline{H} = \underline{B} + \underline{H}$.

Axiom 1 (Eq. 34) gives $\underline{H} + \underline{A} = \underline{H} + \underline{B}$ directly, and so theorem 1 has been proved. The proof of theorem 2 is similar. Now, theorem 3 will be established.

Take p and q as before, and then, since $p = q$, we have $\underline{H}[p] = \underline{H}[q]$ or $\underline{H}[\underline{A}[x]] = \underline{H}[\underline{B}[x]]$, and so $\underline{H} * \underline{A} = \underline{H} * \underline{B}$. Hence, Eq. 246 has been proved. The proof of Eq. 247 is similar.

THEOREM 4. If $\underline{A}$ and $\underline{B}$ are known systems, and it is desired to find an $\underline{H}$ with the property that

$\underline{H} * \underline{A} = \underline{B}$ (248)

then $\underline{H}$, if it exists, is unique.

In other words, there is one and only one system $\underline{H}$ that satisfies Eq. 248. (Of course, there may be no such system. For example, if $\underline{B} = \underline{I}$ and $\underline{A} = \underline{N}_2$, then, because we cannot tell whether x^2 is due to x or $-x$, no $\underline{H}$ exists.) It is assumed that $D_A = D_B$, and $\underline{H}$ is only defined with a domain D_H that equals the range R_A.

To prove this theorem, take $\underline{H}$ and $(\underline{H}+\underline{K})$ to be two systems that satisfy Eq. 248. Then $\underline{H} * \underline{A} = \underline{B}$, and $(\underline{H}+\underline{K}) * \underline{A} = \underline{B}$. Hence, by theorem 1 (Eq. 242), $\underline{H} * \underline{A} - (\underline{H}+\underline{K}) * \underline{A} = \underline{B} - \underline{B}$, or $\underline{K} * \underline{A} = 0$, or $\underline{K}[y] = 0$, with $y = A[x]$ and $x \in D_A$. Therefore, by the definition of the zero system (see sec. 2.5), $\underline{K} = 0$, for domain D_K equal to the range R_A. Hence the system $\underline{H}$ is unique, in this domain.

THEOREM 5. If $\underline{A}$ is a known system, and we desire to find an $\underline{H}$ which is such that

$\underline{A} * \underline{H} = \underline{I}$ (249)

then $\underline{H}$, if it exists, is unique.

To prove this theorem, we precascade Eq. 249 with $\underline{H}$ to obtain $\underline{H} * \underline{A} * \underline{H} = \underline{H}$. By theorem 3 (Eq. 246), this operation is valid. Then we have

$$(\underline{H}*\underline{A}) * \underline{H} = \underline{H} \tag{250}$$

by axiom 5 (Eq. 38). But, Eq. 250 is obviously solved by

$$\underline{H} * \underline{A} = \underline{I} \tag{251}$$

and by theorem 4, Eq. 251 is unique. Now, application of theorem 4 to Eq. 251 shows that $\underline{H}$, if it exists, is unique. Hence theorem 5 is proved.

THEOREM 6. If $\underline{A} * \underline{H} = \underline{I}$ (252)

then $\underline{H} * \underline{A} = \underline{I}$

That $\underline{A} * \underline{H} = \underline{I}$ implies $\underline{H} * \underline{A} = \underline{I}$ was shown in the proof of theorem 5.

THEOREM 7. If $\underline{A}$ and $\underline{B}$ are known systems, and we desire to find an $\underline{H}$ with the property that

$$\underline{H} + \underline{A} = \underline{B} \tag{253}$$

then $$\underline{H} = \underline{B} - \underline{A} \tag{254}$$

uniquely.

Substitution of Eq. 254 in Eq. 253 gives $\underline{B} - \underline{A} + \underline{A} = \underline{B}$, and so $\underline{H} = \underline{B} - \underline{A}$ is a solution. To demonstrate uniqueness, two solutions are assumed and we use the same procedure as in the proof of theorem 4.

6.3 FEEDBACK AND INVERSES

The feedback system that will be investigated is shown in Fig. 28a. This is a sufficiently general system because, as was shown in section 2.8, any single-loop feedback system can be reduced to this form, followed by a feed-through system.

The system equation is $\underline{L} = \underline{I} + \underline{H} * \underline{L}$ or after rearrangement,

$$(\underline{I}-\underline{H}) * \underline{L} = \underline{I} \tag{255}$$

This is recognized as an equation for $\underline{L}$ of the same form as Eq. 249. From theorem 5 we know that $\underline{L}$ is unique. It exists, at least in some sense, because $\underline{L}$ is the feedback system and can be built. Furthermore, from theorem 6, we know that

$$\underline{L} * (\underline{I}-\underline{H}) = \underline{I} \tag{256}$$

Now, if

$$\underline{H} = \underline{I} - \underline{K} \tag{257}$$

then, by substituting Eq. 257 in Eqs. 255 and 256, we obtain

$$\underline{K} * \underline{L} = \underline{I} = \underline{L} * \underline{K} \tag{258}$$

Thus we have shown that it is possible to use a feedback system to construct the unique inverse of a nonlinear system $\underline{K}$. An inverse $\underline{K}^{-1}$ of a system $\underline{K}$ satisfies the property

$$\underline{K}^{-1} * \underline{K} = \underline{K} * \underline{K}^{-1} = \underline{I}$$

The feedback system of Fig. 28a is also an inversion system because we have shown that

$$(\underline{I}-\underline{H}) * \underline{L} = \underline{L} * (\underline{I}-\underline{H}) = \underline{I}$$

Hence the inversion problem and the feedback problem are essentially the same. Therefore, we can write $\underline{L} = \underline{K}^{-1}$.

As it stands, this feedback system for obtaining an inverse (see Fig. 26b) is not a practical physical system because of the unity feedback. A possible way to overcome

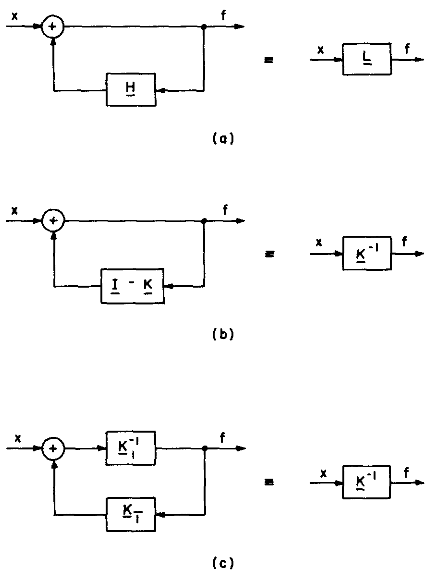


Fig. 28. Feedback and inversion: (a) feedback system; (b) inversion system; (c) equivalent inversion system.

this difficulty is to use the results of section 4.1 to produce the equivalent system shown in Fig. 28b, where $\underline{K}_1^{-1}$ is the inverse of the linear part of $\underline{K}$, that is, $\underline{K}_1$, and $\underline{K}_{\overline{I}}$ form the system $(\underline{K}-\underline{K}_1)$.

The fact that the feedback system defines an algebraically unique system, and that the inversion system produces the unique inverse, does not exhaust the uniqueness problem. Another uniqueness problem will be discussed in the next section.

6.4 INPUT-OUTPUT UNIQUENESS

It is quite easy to visualize a nonlinear system in which the same output is produced by two different inputs. A simple example is found in the no-memory squarer. If we could construct an inverting feedback system of the form of Fig. 28b for such a system, there would be two possible outputs for a single input. Physically, such a situation is untenable, and the inversion system would exhibit some sort of erratic behavior.

This situation of two or more possible outputs for a single input is not limited to inversion systems because the inversion and feedback problems are essentially the same, and hence a feedback system may also exhibit this difficulty.

No-memory feedback or inversion systems are easily handled because the output can be plotted as a function of the input. Therefore any lack of uniqueness at the output is readily detected. In the feedback system of Fig. 28a the impulse responses are bounded functions. We shall show that for such systems the input-output relation must be unique, if an output exists.

We shall consider the system of Fig. 28a in the particular situation $\underline{H} = \underline{H}_1 + \underline{H}_2$, and we shall briefly outline a technique for handling it that was described by Volterra (1). Let the input be x(t), and let there be two possible outputs, f(t) and g(t). Then

$$f = x + \underline{H}[f]$$

and

$$g = x + \underline{H}[g]$$

or

$$f = x + \underline{H}_1[f] + \underline{H}_2[f] \tag{259}$$

and

$$g = x + \underline{H}_1[g] + \underline{H}_2[g] \tag{260}$$

Subtracting Eq. 259 from Eq. 260 yields

$$f - g = \underline{H}_1[f-g] + \underline{H}_2[f] - \underline{H}_2[g] \tag{261}$$

But

$$\begin{aligned}\underline{H}_2[f] - \underline{H}_2[g] &= \underline{H}_2(f^2) - \underline{H}_2(g^2)\\ &= \underline{H}_2(f^2-g^2)\\ &= \underline{H}_2((f+g)(f-g))\end{aligned}$$

Thus Eq. 261 becomes

$$f - g = \underline{H}_1[f-g] + \underline{H}_2((f+g)(f-g)) \tag{262}$$

or

$$p = \underline{H}_1[p] + \underline{H}_2((f+g)p) \tag{263}$$

where p = f - g. Equation 263 can also be written

$$p(t) = \int h_1(\tau)\, p(t-\tau)\, d\tau + \iint h_2(\tau_1, \tau_2)\{f(t-\tau_1)+g(t-\tau_1)\}\, p(t-\tau_2)\, d\tau_1 d\tau_2 \tag{264}$$

where $h_1(t)$ and $h_2(t_1, t_2)$ are the impulse responses of $\underline{H}_1$ and $\underline{H}_2$, respectively.

Now, define

$$k(t, \tau) = h_1(\tau) = \int h_2(\tau_1, \tau)\{f(t-\tau_1) + g(t-\tau_1)\}\, d\tau_1 \tag{265}$$

and Eq. 264 becomes

$$p(t) = \int k(t, \tau)\, p(t-\tau)\, d\tau \tag{266}$$

Since $h_1(t)$ and $h_2(t_1, t_2)$ are bounded functions and $f(t)$ and $g(t)$ are assumed to exist (that is, to be bounded), $k(t, \tau)$ is a bounded function. Then, Eq. 266 can be shown to have a unique solution; that is, $p(t) = 0$. Therefore, $f(t) = g(t)$, and $f(t)$, the system output, is unique.

This can be extended to the situation $\underline{H} = \underline{H}_1 + \underline{H}_2 + \dots + \underline{H}_n$, where the impulse responses are bounded, with the result that the output is unique.

Certain unbounded impulse responses can also be considered by this method. For example, if $\underline{H} = \underline{H}_2 = \underline{A}_1 * \underline{N}_2$, then it can be shown that the impulse response is $h_2(t_1, t_2) = a_1(t_1)\, \delta(t_1-t_2)$, where $a_1(t)$ is the impulse response of $\underline{A}_1$, and $n_2 = 1$. This is an unbounded impulse response, and this case can be shown to be unique. The one place where this test can fail is with $\underline{H} = \underline{N} + \underline{K}$, where $\underline{N}$ is no-memory, and $\underline{K}$ has memory or is zero. This case can exhibit a nonunique input-output relation. It seems to be a fairly safe assumption that this is the only nonunique situation. In case of doubt, the test procedure outlined above can be used to test for uniqueness.

It should be noted that the fact that the system output is unique does not guarantee that the system is well behaved. The output of the system may become unbounded (fail to exist), or some other instability, such as a limit cycle, may exist.

6.5 FUNCTIONAL TAYLOR'S SERIES

We have mentioned that the functional series is closely related to Taylor's series. Here, this relation will be specified in more detail.

Consider a system $\underline{H}$ with an input $ax(t)$ starting at $t = 0$, where a is a constant. Let $f(t_1)$ be the output at a time $t = t_1$. The output will depend on $\underline{a}$; therefore let t_1 be implicit, and consider f as a function of $\underline{a}$ and write it as $f(a)$.

Now we can expand $f(a)$ in a Taylor's series about $a = 0$. Thus

$$f(a) = f(0) + af'(0) + \frac{a^2}{2!} f''(0) + \dots + \frac{a^n}{n!} f^{(n)}(0) + \frac{a^{n+1}}{(n+1)} f^{(n+1)}(\theta) \tag{267}$$

where $f^{(n)}(a)$ is the n^{th} derivative of $f(a)$, and θ is some number between 0 and a.

In particular, if the input is x(t), then a = 1, and we have

$$f(1) = f(0) + f'(0) + \frac{1}{2!} f''(0) + \dots + \frac{1}{n!} f^{(n)}(0) + \frac{1}{(n+1)!} f^{(n+1)}(\theta) \tag{268}$$

Since the input is 0 when a = 0, this can be called a "Taylor's series about zero input." The last term in the series is the error term and if it could be estimated, an idea of the truncation error for Taylor's series could be obtained. Unfortunately, we have not been able to estimate this term.

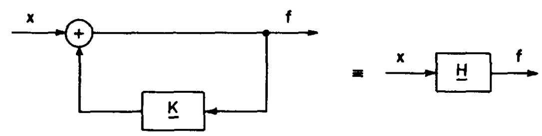


Fig. 29. Feedback system.

In order to illustrate that Eq. 268 is indeed the functional series that constitutes the basis of this report, we shall consider the feedback system of Fig. 29. Then $f = x + \underline{K}[f] = \underline{H}[x]$, and

$$f(a) = ax + \underline{K}_1[f(a)] + \underline{K}_2[f(a)] + \underline{K}_3[f(a)] + \dots \tag{269}$$

Therefore

$$f(0) = 0 + \underline{K}_1 f(0) + \dots \tag{270}$$

Since Eq. 270 is the feedback system with zero input, f(0) = 0. Now, as we know from linear theory,

$$\frac{d}{da} \underline{K}_1[f(a)] = \underline{K}_1[f'(a)]$$

Also, by symmetry,

$$\frac{d}{da} \underline{K}_2[f(a)] = \frac{d}{da} K_2(f^2(a))$$

$$= \frac{d}{da} \iint k_2(\tau_1, \tau_2)\, f(a, t-\tau_1)\, f(a, t-\tau_2)\, d\tau_1 d\tau_2$$

$$= \iint k_2(\tau_1, \tau_2) \frac{d}{da}\{f(a, t-\tau_1) f(a, t-\tau_2)\}\, d\tau_1 d\tau_2$$

$$= \iint k_2(\tau_1, \tau_2) \left\{ \frac{d}{da} f(a, t-\tau_1)\, f(a, t-\tau_2) + f(a, t-\tau_1) \frac{d}{da} f(a, t-\tau_2) \right\} d\tau_1 d\tau_2$$

$$= \underline{K}_2(f'(a)f(a)+f(a)f'(a))$$

$$= \underline{2K}_2(F'(a)f(a))$$

Similarly,

$$\frac{d}{da}\underline{K}_3[f(a)] = \underline{K}_3\left(\frac{d}{da} f^3(a)\right)$$

$$= 3\underline{K}_3(f'(a)f^2(a))$$

and so on, for the higher-order terms. Then

$$f'(a) = x + \underline{K}_1[f'(a)] + 2\underline{K}_2(f'(a)f(a)) + \ldots$$

and

$$f'(0) = x + \underline{K}_1[f'(0)] + 2\underline{K}_2(f'(0)f(0)) + \ldots$$

But $f(0) = 0$, and so $f'(0) = x + \underline{K}_1[f'(0)]$ and, after rearrangement, $f'(0) = (\underline{I}-\underline{K}_1)^{-1}[x]$.
Similarly,

$$f''(0) = \left(2(\underline{I}-\underline{K}_1)^{-1} * \underline{K}_2 \circ \left((\underline{I}-\underline{K}_1)^{-1}\right)^2\right)[x]$$

and so on. Substitution of $f'(0)$, $f''(0)$, etc. in Eq. 268 shows that the resultant series is the same as that obtained in section 2.8 (see Fig. 14).

Since the functional series is so closely related to Taylor's series, we should not expect that the functional series would always converge. The functional series

$$\underline{H} = \underline{H}_1 + \underline{H}_2 + \ldots + \underline{H}_n + \ldots$$

converges for an input $x(t)$ if the output series

$$f(t) = f_1(t) + f_2(t) + \ldots + f_n(t) + \ldots$$

where $f_n(t) = \underline{H}_n[x(t)]$, converges.

For example, consider the system of Fig. 30. The system equation is

$$f(t) = x(t) - \int_{-\infty}^{t} f^2(\tau)\, d\tau \tag{271}$$

When $x(t)$ is a step function, starting at $t = 0$ and with amplitude $+1$, it can be shown that $f(t) = 1/(1+t)$, for $t \geq 0$, by solution of the differential equation (Eq. 271). If the functional series is developed for this series and the result of a unit positive step input is obtained, we find that

$$f(t) = 1 - t + t^2 - t^3 + \ldots \qquad \text{for } t \geq 0$$

This series is not convergent for $t > 1$, but the solution of the differential equation shows that the output is well behaved. Therefore, if the output of a system, which is analyzed by the functional series, is not convergent, we still cannot assume that the output of the system exhibits erratic behavior or becomes unbounded.

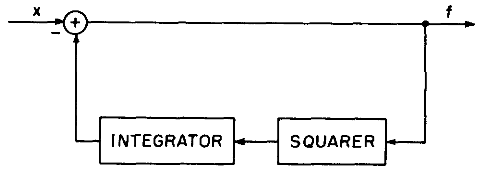


Fig. 30. Simple feedback system.

Brilliant (4) studied the convergence of the functional series and obtained a conservative bounding procedure. His results can be extended somewhat by replacing his norm, $\|k_n\|$, where

$$\|k_n\| = \int_0^\infty \dots \int_0^\infty |k_n(\tau_1, \dots, \tau_n)| \, d\tau_1 \dots d\tau_n$$

by a new norm, $\|k_n\|_T$, where

$$\|k_n\|_T = \int_0^T \dots \int_0^T |k_n(\tau_1, \dots, \tau_n)| \, d\tau_1 \dots d\tau_n$$

In this case we must constrain our considerations of the output to the time interval $0 \leq t \leq T$, rather than to the interval $0 \leq t \leq \infty$ considered by Brilliant. (We assume here that all inputs are zero before $t = 0$, and that the impulse responses are realizable (zero before $t < 0$).)

At the present time, it appears that any general convergence test should be conservative, and that it is best to consider the convergence of each particular case independently. As the example illustrated, convergence difficulties can arise. From a practical point of view, however, the representation is very unwieldy if the convergence is not fairly rapid. But the rapidity of convergence can usually be easily determined in any particular problem by writing out a few terms.

6.6 THE ITERATION SERIES

Even if the functional series that we have used in this report fails to converge, it does not mean that the functional representation fails. There is always the possibility of finding more general functional series that will converge. In this section such a series will be briefly discussed. This series will be called the "iteration series" because it is formed by an iteration procedure.

Consider the feedback system of Fig. 31, with $f = x + \underline{H}[f] = \underline{L}[x]$. A first estimate, $f_{(1)}$, could be made for f, where

$$f_{(1)} = x \tag{272}$$

A second estimate is

$$f_{(2)} = x + \underline{H}[f_{(1)}] = x + \underline{H}[x] \tag{273}$$

and a third estimate is

$$f_{(3)} = x + \underline{H}[f_{(2)}]$$

$$= x + \underline{H}[x+\underline{H}[x]] \tag{274}$$

In general, the n^{th} estimate is

$$f_{(n)} = x + \underline{H}[f_{(n-1)}] \tag{275}$$

If we let

$$f_{(n)} = \underline{L}_{(n)}[x] \tag{276}$$

then $\underline{L}_{(n)}$ is an approximation to the actual system $\underline{L}$, and

$$\underline{L}_{(1)} = \underline{I}$$

$$\underline{L}_{(2)} = \underline{I} + \underline{H}$$

$$\underline{L}_{(3)} = \underline{I} + \underline{H} * (\underline{I}+\underline{H})$$

by applying Eq. 276 to Eqs. 272, 273, and 274. In general,

$$\underline{L}_{(n)} = \underline{I} + \underline{H} * \underline{L}_{(n-1)} \tag{277}$$

In the limit as $n \to \infty$, direct substitution of $\underline{L}_{(n)}$ in the system equation, $\underline{L} = \underline{I} + \underline{H} * \underline{L}$, for Fig. 31, shows that the equation is satisfied. In practice $\underline{L}_{(n)}$ would not be used in the limit, but would be truncated at some point; that is, $\underline{L}_{(n)}$ would be used with a finite n as an approximation to $\underline{L}$.

Any physical feedback system has a delay around the loop. This delay is usually too small to be important, but it has an interesting effect on the iteration series. Let the feedback element $\underline{H}$ be replaced by $\underline{D} * \underline{H}$, where $\underline{D}$ is a delay. If this is done, it can be shown that the iteration series is automatically truncated at some $\underline{L}_{(n)}$, where n depends on the length of the delay and on the time after the input has started at which the output is being observed. This occurs because the iteration procedure of Eqs. 272-275, with $\underline{H}$ replaced by $\underline{D} * \underline{H}$, is the actual sequence of operations in the system. In the first time interval, 0 to δ, where δ is the delay time, the output is x(t) because the delay holds back the feedback. In the next interval, δ to 2δ, the output is $x(t) + \underline{D} * \underline{H}[x(t)]$, and so on for each interval. This is precisely the iteration

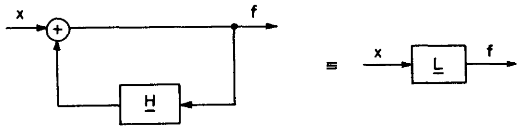


Fig. 31. Feedback system.

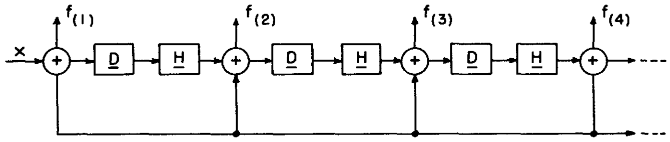


Fig. 32. Diagram of iteration series.

procedure. This can also be visualized by means of Fig. 32 (which was suggested by G. D. Zames). This system is equivalent to the feedback system of Fig. 31, with $\underline{H}$ replaced by $\underline{D} * \underline{H}$.

The iteration series, then, is closely tied in with the actual sequence of operations in a feedback system. Therefore, it seems reasonable to assume that if the iteration series tends to become unbounded, the actual system output will tend to become unbounded. (We note that the delay in a physical system will keep the output bounded at all finite times, but it could get larger as time increases and the system would then be said to be unstable.) A very simple example of this is the system of Fig. 31 when $\underline{H} = A\,\underline{I}$, where A is a constant. Using the iteration series for this system yields

$$f(t) = x(t)(1+A+A^2+A^3+\dots) \tag{278}$$

and this series becomes unbounded if $A > 1$. It is known from linear theory that the system is unstable if $A > 1$.

Thus, we see that the iteration series is much more closely connected with the physical world then the functional Taylor's series. As a practical tool it is not, now, very useful because experience has shown that, when the iteration series is rapidly convergent, the functional Taylor's series is also rapidly convergent. However, it does present the possibility of using other functional series than the functional Taylor's series.

The convergence of the iteration series can often be determined. If a linear feedback system (Fig. 31 with $\underline{H} = \underline{H}_1$) has an input that starts at $t = 0$, and the impulse response $h_1(t)$ is bounded, then at any finite time t, the iteration series can be shown to be convergent (11).

Let us assume for a feedback system (Fig. 31) with a nonlinear feedback element $\underline{H}$ that the following (Lipschitz) condition holds:

$$\left|\underline{H}[x]-\underline{H}[y]\right| \leq \left|\underline{K}_1[x-y]\right| \tag{279}$$

where x and y are any input signals, and $\underline{K}_1$ is some linear system with a bounded impulse response. (Actually, the impulse response $k_1(t)$ need be bounded only over a time interval 0 to T if we limit our consideration of the output to this interval.) Then we can show that the iteration series for this nonlinear system converges at any finite time after the input has started. This is done by appropriately bounding the terms of the iteration series of the nonlinear system by the terms of the iteration series of the linear system obtained by replacing $\underline{H}$ by $\underline{K}_1$. Then, since the iteration series for the linear system is known to converge, the iteration series of the nonlinear system will converge. Referring to the results of Section V, we recall that the output of such a system is unique, and hence the iteration series converges to the true output of the nonlinear feedback system.

A system $\underline{H}$ that satisfies Eq. 279 might well be called a "saturation" system because it is bounded by a linear system. An ordinary saturation curve (for example, the flux-current characteristic of a magnetic material) is bounded by a linear curve.

6.7 CONTINUOUS SYSTEMS

Let us consider a system $\underline{H}$ with an input x(t) and an input y(t). From an intuitive point of view, we could say that $\underline{H}$ is continuous if $\underline{H}[x(t)]$ and $\underline{H}[y(t)]$ are close together and if x(t) and y(t) are close together.

Brilliant (4) defined a much more rigorous concept of continuity. First, define a distance between input functions x(t) and y(t):

$$\left| \int_{t-r}^{t} \{x(\tau)-y(\tau)\} \, d\tau \right|$$

for $r > 0$. Define another distance:

$$\left| \underline{H}[x(t)] - \underline{H}[y(t)] \right|$$

between the outputs $\underline{H}[x(t)]$ and $\underline{H}[y(t)]$. Now we have a precise measurement of distance, and so closeness and continuity can be rigorously defined. The following definition is not the only possible definition of continuity, nor are these distances the only possible distances that could be defined.

If we have a time-stationary system $\underline{H}$, and bounded inputs x(t) and y(t), Brilliant's definition of continuity can be stated: $\underline{H}$ is continuous if for any $\epsilon > 0$, there exists a $T > 0$, $\delta > 0$ (T sufficiently large, δ sufficiently small), such that, if $\left| \int_{t-r}^{t} \{x(\tau) - y(\tau)\} \, d\tau \right| < \delta$, for $0 \leq r \leq T$, then $\left| \underline{H}[x(t)] - \underline{H}[y(t)] \right| < \epsilon$. Brilliant also showed that if $\underline{H}$ is continuous, then for any $\epsilon > 0$ there is a polynomial system $\underline{H}_{(\epsilon)}$, consisting of the sum of a constant, a linear system with Lebesque integrable impulse response, and products of such linear systems, such that, for any bounded input x(t), $\left| \underline{H}[x(t)] - \underline{H}_{(\epsilon)}[x(t)] \right| < \epsilon$. That is to say, if $\underline{H}$ is continuous, then it can be closely approximated by

$$\underline{H}_0 + \underline{H}_1 + \dots + \underline{H}_n$$

where

$$\underline{H}_i = \sum \underbrace{\underline{A}_1 \cdot \underline{B}_1 \cdot \dots \cdot \underline{M}_1}_{i \text{ systems}}$$

The sum is over a number of such products.

This is a sufficient condition only for a system $\underline{H}$ to be approximated by the functional representation. Certainly other systems can be approximated. The statement is, however, a precise mathematical theorem describing a set of systems that can be expanded in the functional representation.

6.8 CONCLUSION

In conclusion, it is appropriate to mention some of the future prospects for the functional analysis of nonlinear systems.

First, of course, there is the application of this form of analysis to actual engineering problems. Moreover, the general nature of this system representation makes it a possible tool in the investigation of several general problems, for example, the question of what constraints on the open-loop response of a nonlinear feedback system are necessary to ensure the stability of the closed loop.

The functional representation, as it now stands, still presents problems. The principal problem is that of obtaining a series that has rapid convergence when the functional Taylor's series is not rapidly convergent, or not convergent at all.

A situation of considerable interest occurs when non-Gaussian random signals are being investigated. Optimization problems then point to nonlinear systems, and the functional representation seems to be a good system description to use for such problems.

APPENDIX A

TRANSFORMS

1. SYSTEM TRANSFORMS

The one-dimensional Fourier-transform pair is defined as

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt \tag{A-1}$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega)\, e^{+j\omega t}\, d\omega \tag{A-2}$$

A multidimensional generalization of this is the transform pair:

$$F_n(j\omega_1, \ldots, j\omega_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_n(t_1, \ldots, t_n)\, dt_1 \ldots dt_n \, \exp(-j\omega_1 t_1 - \ldots - j\omega_n t_n) \tag{A-3}$$

$$f_n(t_1, \ldots, t_n) = \left(\frac{1}{2\pi}\right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F_n(j\omega_1, \ldots, j\omega_n) \exp(+j\omega_1 t_1 + \ldots + j\omega_n t_n)\, d\omega_1 \ldots d\omega_n \tag{A-4}$$

Another generalized relation is the multidimensional Parseval theorem (6):

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_n(t_1, \ldots, t_n)\, g_n(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

$$= \left(\frac{1}{2\pi}\right)^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F_n(j\omega_1, \ldots, j\omega_n)\, G_n(j\omega_1, \ldots, j\omega_n)\, d\omega_1 \ldots d\omega_n \tag{A-5}$$

where $f_n(t_1, \ldots, t_n)$ and $F_n(j\omega_1, \ldots, j\omega_n)$, and $g_n(t_1, \ldots, t_n)$ and $G_n(j\omega_1, \ldots, j\omega_n)$ are Fourier-transform pairs.

If $f(t) = 0$ for $t < 0$ and

$$\int_{0}^{\infty} |f(t)|\, e^{-\sigma t}\, dt < \infty \tag{A-6}$$

where σ is a real number, then we can define the unilateral Laplace-transform pair:

$$F(s) = \int_{0}^{\infty} f(t)\, e^{-st}\, dt \tag{A-7}$$

$$f(t) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} F(s)\, e^{+st}\, ds \tag{A-8}$$

where $s = \sigma + j\omega$. The multidimensional generalization of this is

$$F_n(s_1, \ldots, s_n) = \int_0^{\infty} \cdots \int_0^{\infty} f_n(t_1, \ldots, t_n)\, dt_1 \ldots dt_n \exp(-s_1t_1 - \ldots - s_nt_n) \tag{A-9}$$

$$f_n(t_1, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^n \int_{\sigma_1 - j\infty}^{\sigma_1 + j\infty} \cdots \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} F_n(s_1, \ldots, s_n) \exp(+s_1t_1 + \ldots + s_nt_n)\, ds_1 \ldots ds_n \tag{A-10}$$

where $\sigma_1, \ldots, \sigma_n$ are real numbers. If $f_n(t_1, \ldots, t_n)$ is symmetrical, then $\sigma_1 = \ldots = \sigma_n = \sigma$.

2. ASSOCIATION OF VARIABLES

Assume that we are given $f_n(t_1, \ldots, t_n)$ and its transform $F_n(s_1, \ldots, s_n)$. This transform may be Fourier ($s_i = j\omega_i$), or Laplace. The problem is to find the transform of $f_n(t_1, t_2, t_3, \ldots, t_n)$ from $F_n(s_1, \ldots, s_n)$. (Actually, we are associating any two variables t_i and t_j. For convenience, take $i = 1$, $j = 2$. There is no loss in generality.) Call this transform $G_{n-1}(s_1, s_3, \ldots, s_n)$. Now, by setting $t_2 = t_1$ in Eq. A-10, we have

$$f_n(t_1, t_1, t_3, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^n \int_{\sigma_1 - j\infty}^{\sigma_1 + j\infty} \cdots \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} F_n(s_1, \ldots, s_n) \exp(+s_1t_1 + s_2t_1 + s_3t_3 + \ldots + s_nt_n)\, ds_1 \ldots ds_n \tag{A-11}$$

Then

$$f_n(t_1, t_1, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^n \int_{\sigma_2 - j\infty}^{\sigma_2 + j\infty} \cdots \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} \left\{ \int_{\sigma_1 - j\infty}^{\sigma_1 + j\infty} F_n(s_1, \ldots, s_n) \exp(+(s_1+s_2)t_1)\, ds_1 \right\} \exp(+s_3t_3 + \ldots + s_nt_n)\, ds_2 \ldots ds_n \tag{A-12}$$

Setting $s_1 + s_2 = u_1$ gives

$$f_n(t_1, t_1, t_3, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^n \int_{\sigma_2 - j\infty}^{\sigma_2 + j\infty} \cdots \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} \left\{ \int_{\sigma - j\infty}^{\sigma + j\infty} F_n(u_1 - s_2, s_2, s_3, \ldots, s_n) \exp(+u_1t_1)\, du_1 \right\} \exp(+s_3t_3 + \ldots + s_nt_n)\, ds_2 \ldots ds_n$$

$$= \left(\frac{1}{2\pi j}\right)^{n-1} \int_{\sigma - j\infty}^{\sigma + j\infty} \int_{\sigma_3 - j\infty}^{\sigma_3 + j\infty} \cdots \int_{\sigma_n - j\infty}^{\sigma_n + j\infty} \left\{ \frac{1}{2\pi j} \int_{\sigma_2 - j\infty}^{\sigma_2 + j\infty} F_n(u_1 - s_2, s_2, s_3, \ldots, s_n)\, ds_2 \right\} du_1 ds_3 \ldots ds_n \tag{A-13}$$

where σ is chosen so that the integral converges. Now

$$f_n(t_1, t_1, t_3, \ldots, t_n) = \left(\frac{1}{2\pi j}\right)^{n-1} \int_{\sigma-j\infty}^{\sigma+j\infty} \int_{\sigma_3-j\infty}^{\sigma_3+j\infty} \cdots \int_{\sigma_n-j\infty}^{\sigma_n+j\infty}$$

$$G_{n-1}(s_1, s_3, \ldots, s_n) \exp(+s_1t_1+s_3t_3+\ldots+s_nt_n)\, dt_1 dt_3 \ldots dt_n \tag{A-14}$$

and hence, if s_2 is replaced by u, and u_1 by s_1, in Eq. A-13, we have

$$G_{n-1}(s_1, s_3, \ldots, s_n) = \frac{1}{2\pi j} \int_{\sigma_2-j\infty}^{\sigma_2+j\infty} F_n(s_1-u, u, s_3, \ldots, s_n)\, du \tag{A-15}$$

Specification of n = 2 gives Eq. 115.

3. FINAL-VALUE AND INITIAL-VALUE THEOREMS

Consider a multidimensional function, $f_n(t_1, \ldots, t_n)$, and its transform, $F_n(s_1, \ldots, s_n)$. Define

$$g_1(t_1; s_2, \ldots, s_n) = \int_0^\infty \cdots \int_0^\infty f_n(t_1, \ldots, t_n) \exp(-s_2t_2-\ldots-s_nt_n)\, dt_2 \ldots dt_n \tag{A-16}$$

Then the first-order transform of $g_1(t_1; s_2, \ldots, s_n)$ is

$$G_1(s_1; s_2, \ldots, s_n) = \int_0^\infty g_1(t_1; s_2, \ldots, s_n) \exp(-s_1t_1)\, dt_1$$

$$= F_n(s_1, \ldots, s_n) \tag{A-17}$$

Now, if we regard $s_2, \ldots, s_n$ as fixed, $g_1(t_1; s_2, \ldots, s_n)$ and $G_1(s_1; s_2, \ldots, s_n)$ make a first-order transform pair. Then, from linear theory (9), we have

$$\lim_{t_1\to\infty} g_1(t_1; s_2, \ldots, s_n) = \lim_{s_1\to 0} G_1(s_1; s_2, \ldots, s_n)\, s_1 \tag{A-18}$$

and therefore

$$\lim_{t_1\to\infty} g_1(t_1; s_2, \ldots, s_n) = \lim_{s_1\to 0} F_n(s_1, \ldots, s_n)\, s_1 \tag{A-19}$$

But

$$\lim_{t_1\to\infty} g_1(t_1; s_2, \ldots, s_n) = \lim_{t_1\to\infty} \int_0^\infty \cdots \int_0^\infty f_n(t_1, \ldots, t_n) \exp(-s_2t_2-\ldots-s_nt_n)\, dt_2 \ldots dt_n$$

$$= \int_0^\infty \cdots \int_0^\infty \left\{ \lim_{t_1\to\infty} f_n(t_1, \ldots, t_n) \right\} \exp(-s_2t_2-\ldots-s_nt_n)\, dt_2 \ldots dt_n \tag{A-20}$$

Therefore the transform of $\lim\limits_{t_1 \to \infty} f_n(t_1, \ldots, t_n)$ is

$$\lim_{s_1 \to 0} F_n(s_1, \ldots, s_n)\, s_1 \tag{A-21}$$

This justifies the limiting procedure used in section 5.4.

Successive use of Eq. A-21 shows that

$$\lim_{\substack{t_1 \to \infty \\ \vdots \\ t_n \to \infty}} f_n(t_1, \ldots, t_n) = \lim_{\substack{s_1 \to 0 \\ \vdots \\ s_n \to 0}} F_n(s_1, \ldots, s_n)\, s_1 \ldots s_n \tag{A-22}$$

Now

$$\lim_{\substack{t_1 \to \infty \\ \vdots \\ t_n \to \infty}} f_{(n)}(t_1, \ldots, t_n) = \lim_{t \to \infty} f_n(t) \tag{A-23}$$

where $f_n(t) = f_{(n)}(t, t, \ldots, t)$, and so the final-value theorem of section 3.4 is proved.

In a similar manner, the initial-value theorem of section 3.4 can be proved.

APPENDIX B

PROBLEM DETAILS

In this appendix, we shall give further details for some of the problems of Sections III and V.

1. PROBLEMS OF SECTION 3.5

System A is given by

$$\underline{K} = \underline{N} * \underline{H}_1$$

$$= \underline{H}_1 + n_3 \underline{H}_1^3 + n_5 \underline{H}_1^5$$

$$= \underline{K}_1 + \underline{K}_3 + \underline{K}_5 \tag{B-1}$$

where the coefficients associated with the nonlinear no-memory system $\underline{N} = \underline{I} + \underline{N}_3 + \underline{N}_5$ are n_3 and n_5. Since $\underline{H}_1$ has a transform $H/(s+\alpha)$, the transforms of the system $\underline{K}$ are

$$\underline{K}_1(s) = \frac{H}{s+\alpha} \tag{B-2}$$

$$K_3(s_1, s_2, s_3) = \frac{n_3 H^3}{(s_1+\alpha)(s_2+\alpha)(s_3+\alpha)} \tag{B-3}$$

$$K_5(s_1, \ldots, s_5) = \frac{n_5 H^5}{(s_1+\alpha) \ldots (s_5+\alpha)} \tag{B-4}$$

by use of Eq. 90. The input is $\text{Re}\left\{x^{j\omega t}\right\}$, and, from section 3.3, the complex amplitude of the first-harmonic output is

$$K_1(j\omega) + \frac{3}{4} K_3(j\omega, j\omega, -j\omega) + \frac{5}{8} K_5(j\omega, j\omega, j\omega, -j\omega, -j\omega) \tag{B-5}$$

The third-harmonic complex output amplitude is

$$\frac{1}{4} K_3(j\omega, j\omega, j\omega) + \frac{5}{16} K_5(j\omega, j\omega, j\omega, j\omega, -j\omega) \tag{B-6}$$

and the fifth-harmonic complex output amplitude is

$$\frac{1}{16} K_5(j\omega, \ldots, j\omega) \tag{B-7}$$

Substituting the expressions for the transforms (Eqs. B-2, B-3, and B-4), taking $\omega = 0$, and defining $B = H/\alpha$ (the linear undistorted gain) gives the low-frequency amplitudes:

$$\text{first-harmonic amplitude} = B + \frac{3}{4} n_3 B^3 x^3 + \frac{5}{8} n_5 B^5 x^5 \tag{B-8}$$

$$\text{third-harmonic amplitude} = \frac{1}{4} n_3 B^3 X^3 + \frac{5}{16} n_5 B^5 X^5 \tag{B-9}$$

$$\text{fifth-harmonic amplitude} = \frac{1}{16} n_5 B^5 X^5 \tag{B-10}$$

Defining

$$\text{first-harmonic distortion} = \frac{\text{first-harmonic amplitude} - B}{B} \tag{B-11}$$

$$\text{third-harmonic distortion} = \frac{\text{third-harmonic amplitude}}{B} \tag{B-12}$$

$$\text{fifth-harmonic distortion} = \frac{\text{fifth-harmonic amplitude}}{B} \tag{B-13}$$

and substituting Eqs. B-8 to B-10 gives Eqs. 145-147 of example 2.

Now the feedback system B can be considered. The system equation is

$$\underline{L} = \underline{N} * \underline{H}_1 * (A\underline{L}) \tag{B-14}$$

where A is a gain constant, and $\underline{L}$ represents the feedback system. Substituting Eq. 144 for $\underline{N}$, taking

$$\underline{L} = \underline{L}_1 + \underline{L}_2 + \dots \tag{B-15}$$

and determining $\underline{L}_1$, $\underline{L}_2$, etc. by the methods of section 2.8 yields

$$\underline{L}_1 = \underline{H}_1 * (\underline{I} + A\underline{H}_1)^{-1} \tag{B-16}$$

$$\underline{L}_3 = n_3 (\underline{I} + A\underline{H}_1)^{-1} * \left(\underline{L}_1^3\right) \tag{B-17}$$

$$\underline{L}_5 = (\underline{I} + A\underline{H}_1)^{-1} * \left(3n_3\left(\underline{L}_1^2 \cdot (A\underline{H}_1 * \underline{L}_3)\right) + n_5 \underline{L}_1^5\right) \tag{B-18}$$

for the first three terms. (All even terms are zero.)

The transforms that are found by the relations in section 2.9, are:

$$L_1(s) = \frac{H}{s + a + AH} \tag{B-19}$$

$$L_3(s_1, s_2, s_3) = \frac{s_1 + s_2 + s_3 + AH}{s_1 + s_2 + s_3 + a + AH} \, n_3 \, \frac{H}{s_1 + a + AH} \, \frac{H}{s_2 + a + AH} \, \frac{H}{s_3 + a + AH} \tag{B-20}$$

$$L_5(s_1, \dots, s_5) = \frac{s_1 + \dots + s_3 + a}{s_1 + \dots + s_5 + a + AH} \left\{ 3n_3 L_1(s_1) \, L_1(s_2) \, \frac{AH}{s_3 + s_4 + s_5 + a} \times L_3(s_1 + s_2 + s_3) + n_5 L_1(s_1) \dots L_1(s_5) \right\} \tag{B-21}$$

Calculating the distortion ratios as we did in Eqs. B-8 to B-13 gives Eqs. 148, 149, and 150 of example 2. If the conditions on the size of the distortion ratios given in example 3 are met, then these three terms of the series for $\underline{L}$ will be sufficient.

2. PROBLEMS OF SECTIONS 3.6 AND 3.7

From the transform $\underline{N} = \underline{I} + \underline{N}_3 + \underline{N}_5$, the system equation is

$$\underline{L} = A\underline{N} * \underline{H}_1 * (\underline{I}-\underline{L}) \tag{B-22}$$

where A is the gain constant, $\underline{H}_1$ is an ideal integrator, and

$$\underline{N}[y] = \sin y \tag{B-23}$$

Taking

$$\underline{L} = \underline{L}_1 + \underline{L}_2 + \dots \tag{B-24}$$

and applying the methods of section 2.6 gives

$$\underline{L}_1 = (\underline{I}+A\underline{H}_1) * A\underline{H}_1 \tag{B-25}$$

$$\underline{L}_3 = -(\underline{I}+A\underline{H}_1)^{-1} * \frac{A}{3!}\left(\frac{1}{A}\,\underline{L}_1\right)^3 \tag{B-26}$$

$$\underline{L}_5 = (\underline{I}+A\underline{H}_1)^{-1} * \left(\frac{1}{2}\,A\left(\frac{1}{A}\,\underline{L}_1\right)^3 \cdot \underline{L}_3 + \frac{A}{5!}\left(\frac{1}{A}\,\underline{L}_1\right)^5\right) \tag{B-27}$$

The transforms are found by the relations in section 2.9, and are given in Eqs. 152, 153, and 154 of example 3. The output transforms are given in Eqs. 156, 157, and 158.

The output transform $R_3(s)$ is obtained from Eq. 157, by the association procedure, as follows:

$$R_{(3)}(s_1, s_2, s_3) = -\frac{AS^3}{3!}\,\frac{(s_1+s_2+s_3)}{s_1+s_2+s_3+A}\,\frac{1}{s_1(s_2+A)}\,\frac{1}{s_2(s_2+A)}\,\frac{1}{s_3(s_3+A)}$$

$$= -\frac{AS^3}{6}\,\frac{(s_1+s_2+s_3)}{(s_1+s_2+s_3+A)}\,\frac{1}{s_1(s_1+A)}\,\frac{1}{A^2}\left\{\frac{1}{s_2} - \frac{1}{s_2+A}\right\}\left\{\frac{1}{s_3} - \frac{1}{s_3+A}\right\} \tag{B-28}$$

Associating s_2 and s_3 gives

$$R_3(s) = -\frac{S^3}{6A}\,\frac{s_1+s_2}{s_1+s_2+A}\,\frac{1}{s_1(s_1+A)}\left\{\frac{1}{s_2} - \frac{2}{s_2+A} + \frac{1}{s_2+2A}\right\}$$

$$= -\frac{S^3}{6A^2}\,\frac{s_1+s_2}{s_1+s_2+A}\left\{\frac{1}{s_1} - \frac{1}{s_1+A}\right\}\left\{\frac{1}{s_2} - \frac{2}{s_2+A} + \frac{1}{s_2+2A}\right\} \tag{B-29}$$

and associating s_1 and s_2 gives

$$R_3(s) = -\frac{S^3}{6A^2}\frac{s}{s+A}\left\{\frac{1}{s} - \frac{3}{s+A} - \frac{3}{s+2A} - \frac{1}{s+3A}\right\} \tag{B-30}$$

Now,

$$R_{(5)}(s_1, \ldots, s_5) = -\frac{S^5A^2}{2}\frac{(s_1+\ldots+s_5)}{s_1+\ldots+s_5+A}\frac{1}{s_1+s_2+s_3+A}\frac{1}{s_1(s_1+A)} \cdots \frac{1}{s_5(s_5+A)} \tag{B-31}$$

where the second term with coefficient $\frac{1}{5!}$ has been neglected because it is small. Carrying out the association procedure on this basis gives

$$R_5(s) = \frac{1}{2}\frac{S^5}{A^4}\frac{s}{s+A}\left\{\frac{1}{s} - \frac{0.5}{s+A} - \frac{5}{s+2A} + \frac{8}{s+3A} - \frac{4}{s+4A} + \frac{0.5}{s+5A} - \frac{3A}{(s+A)^2} + \frac{6A}{(s+2A)^2} - \frac{3A}{(s+3A)^2}\right\} \tag{B-32}$$

The transform of the output is

$$R(s) = R_1(s) + R_2(s) + R_3(s) \tag{B-33}$$

and the terms can be collected and inverted by the usual linear methods to give the output $r(t)$ ($r(t)$ is given in Eq. 159 in example 4). Only one of the multiple-order poles gives a significant contribution if $A \gg 1$.

3. PROBLEMS OF SECTION 5.5

The transform of $l_2(t_1, t_2)$ is

$$L_2(s_1, s_2) = \frac{A^2B}{(s_1+s_2+\beta)(s_1+\alpha)(s_2+\alpha)} \tag{B-34}$$

and we want to calculate

$$\int l_2(\tau, \tau)\, d\tau \tag{B-35}$$

by the method of section 5.4. Therefore we associate s_1 and s_2 in Eq. B-34, which gives

$$\frac{A^2B}{(s+\beta)(s+2\alpha)} \tag{B-36}$$

Then

$$\int l_2(\tau,\tau)\, d\tau = \lim_{s\to 0} \frac{A^2B}{(s+\beta)(s+2\alpha)}$$

$$= \frac{A^2B}{2\alpha\beta}$$

$$= \frac{A^2}{2\alpha} \qquad \text{(B-37)}$$

if $B = \beta$, and so the first term of Eq. 218 is

$$\left(\frac{A^2}{2\alpha}\right)^2 \qquad \text{(B-38)}$$

Next, we want to calculate the second term of Eq. 218. Equation 219 is the transform of

$$2\iint l_2(\tau_1, \tau_2)\, l_2(T_1+\tau_1, T_2+\tau_2)\, d\tau_1 d\tau_2 \qquad \text{(B-39)}$$

and we want to associate s_1 and s_2 in Eq. 219 (T_1 and T_2 in Eq. B-39). Thus

$$2L_2(-s_1, -s_2)\, L_2(s_1, s_2) = \frac{2A^4B^2}{(-s_1-s_2+\beta)(s_1+s_2+\beta)}\, \frac{1}{4\alpha^2} \left\{\frac{1}{-s_1+\alpha} + \frac{1}{s_1+\alpha}\right\}$$

$$\left\{\frac{1}{-s_2+\alpha} + \frac{1}{s_2+\alpha}\right\} \qquad \text{(B-40)}$$

Associating s_1 and s_2 gives

$$\frac{A^4B^2}{2\alpha^2}\, \frac{1}{-s+\beta}\, \frac{1}{s+\beta} \left\{\frac{1}{-s+2\alpha} + \frac{1}{s+2\alpha}\right\} = 2A^2B^2 \left(\frac{1}{-s^2+\beta^2}\right)\left(\frac{1}{-s^2+4\alpha^2}\right) \qquad \text{(B-41)}$$

This is the transform of the second term of Eq. 218. Setting $s = j\omega$ and combining the two terms gives the transform of $\phi_f(T)$, which is the spectrum, $\Phi_f(\omega)$. The result is Eq. 220 in example 5.

4. PROBLEMS OF SECTION 5.6

First, we obtain Eq. 222. Linear system analysis shows that

$$E(s) = \frac{s}{s+A}\, X(s) \qquad \text{(B-42)}$$

where $E(s)$ and $X(s)$ are the transforms of $e(t)$ and $x(t)$, respectively. But

$$X(s) = \frac{B}{s+\beta}\, Y(s) \qquad \text{(B-43)}$$

where $Y(s)$ is white Gaussian noise, and $B/(s+\beta)$, from $\Phi_x(\omega)$ in Eq. 221, is the shaping filter. Then

$$e = \underline{K}_1[y] \tag{B-44}$$

where $\underline{K}_1$, from Eqs. B-42 and B-43, has a transform

$$\frac{s}{s+A} \cdot \frac{B}{s+\beta} \tag{B-45}$$

Now

$$\overline{e^2(t)} = \int k_1(\tau)\, k_1(\tau)\, d\tau \tag{B-46}$$

and, by application of the output averages of section 5.2, ($\overline{e^2(t)}$ is $\overline{e(t)\ e(t+T)}$ evaluated at T = 0). The transform of $k_1(t_1)\, k_1(t_2)$ is

$$\frac{Bs_1}{(s_1+A)(s_1+\beta)}\ \frac{Bs_3}{(s_2+A)(s_2+\beta)} \tag{B-47}$$

and by associating s_1 and s_2, the transform of $k_1(t)\, k_1(t_2)$ is

$$\frac{B^2}{(A-\beta)^2}\left\{\frac{A^2}{s+2A} - \frac{2\beta}{s+\beta+A} + \frac{\beta^2}{a+2\beta}\right\} \tag{B-48}$$

Assuming that $A \gg \beta$, we have

$$e^2(t) = \lim_{s \to 0} \frac{B^2}{A^2}\ \frac{A^2}{s+2A}$$

$$= \frac{B^2}{2A} \tag{B-49}$$

which is Eq. 222 in example 6. Equations 223 and 224 are obtained in a similar manner. (Equation B-49 can be obtained by standard linear methods (8); it is derived here by the methods of Section V, in order to illustrate this application to linear analysis.)

The nonlinear compensated system will now be considered. The first two system transforms, $L_1(s)$ and $L_3(s_1, s_2, s_3)$, are given in Eqs. 105 and 106. The output error is given by

$$e(t) = f(t) - x(t) \tag{B-50}$$

where $x(t)$ is related to the white Gaussian $y(t)$ by Eq. B-43. We can show (assuming that n_3 is sufficiently small that only $e_1(t)$ and $e_3(t)$ contribute significantly to the error $e(t)$) that

$$e(t) = e_1(t) + e_3(t) \tag{B-51}$$

where

$$e_1 = \underline{K}_1[y] \tag{B-52}$$

$$e_3 = \underline{K}_3[y] \tag{B-53}$$

and $\underline{K}_1$ and $\underline{K}_3$ have transforms

$$K_1(s) = \frac{B}{s + A} \tag{B-54}$$

$$K_3(s_1, s_2, s_3) = \frac{An_3}{s_1 + s_2 + s_3 + A} \frac{B}{s_1 + A} \frac{B}{s_2 + A} \frac{B}{s_3 + A} \tag{B-55}$$

Here, we have taken $\beta = 0$. Then

$$\overline{e^2(t)} = \overline{e_1^2(t)} + 2\overline{e_1(t)\, e_3(t)} + \overline{e_3^2(t)} \tag{B-56}$$

where $\overline{e_1^2(t)}$ is given in Eq. B-49. The other two terms of $\overline{e^2(t)}$ can be computed by the methods of section 5.4.

To illustrate this point we shall compute $\overline{e_1(t)\, e_2(t)}$. Now

$$\overline{e_1(t)\, e_2(t)} = (\underline{K}_1 \cdot \underline{K}_3)\overline{(y_1 y_2 y_3 y_4)}$$

$$= 3 \iint k_1(\tau_1)\, k_3(\tau_1, \tau_2, \tau_3)\, d\tau_1 d\tau_2 \tag{B-57}$$

from Eq. 199, with $T = 0$. Let

$$\int k_3(t, \tau, \tau)\, d\tau = b(t) \tag{B-58}$$

The transform of this term, B(s), has been worked out in Eqs. 213 and 214-216. Taking $K = AB^3 n_3$ and $\alpha = \beta = A$ in Eq. 216, gives

$$B(s) = \frac{B^3 n_3}{2} \left(\frac{1}{s+A}\right)^2 \tag{B-59}$$

The transform of $k_1(t_1)\, b(t_2)$ is

$$\frac{3B}{s_1 + A} \frac{B^3 n_3}{2} \left(\frac{1}{s_2 + A}\right)^2 \tag{B-60}$$

Now we can complete the evaluation of Eq. B-57 by associating s_1 and s_2 in Eq. B-60 to give

$$\frac{3B^4 n_3}{2} \left(\frac{1}{s+2A}\right)^2 \tag{B-61}$$

Then

$$\overline{l_1(t)\ l_2(t)} = \lim_{s \to 0} \frac{3B^4 n_3}{2} \left(\frac{1}{s+2A}\right)^2$$

$$= \frac{3}{8} \frac{B^4 n_3}{A^2} \qquad \text{(B-62)}$$

In a similar manner, $\overline{e_2^2(t)}$ can be computed.

5. PROBLEMS OF SECTION 5.9

The desired operation on $y(t)$ is a predictor $\underline{P}$, with the desired output

$$d(t) = y(t+T)$$
$$= \underline{P}[y(t)]$$
$$= \int \delta(\tau - T)\, y(t-\tau)\, d\tau \qquad \text{(B-63)}$$

$\underline{P}$, therefore, has an impulse response $p(t) = \delta(t-T)$. The white Gaussian $x(t)$ is obtained by operating on $y(t)$ to produce

$$x = \left(\underline{L}_1^{-1} * \underline{N}^{-1}\right)[y] \qquad \text{(B-64)}$$

whence the desired operation on $x(t)$ to produce $d(t)$ is

$$\underline{P} * (\underline{N} * \underline{L}_1) = \underline{P} * (\underline{N}_1 + \underline{N}_3) * \underline{L}_1$$

$$= \underline{P} * \underline{L}_1 + \underline{P} * \left(\underline{L}_1^3\right) \qquad \text{(B-65)}$$

This operation is the sum of a first-order system and a third-order system, and the impulse responses are

$$l_1(t+T) \qquad \text{(B-66)}$$

and

$$l_1(t_1+T)\ l_2(t_2+T)\ l_3(t_3+T) \qquad \text{(B-67)}$$

where $l_1(t)$ is the impulse response of $\underline{L}_1$.

Inverting $L_1(s)$ in Eq. 233 and applying Eq. 230 gives a best realizable system $\underline{H}$, with

$$\underline{H} = \underline{H}_{(1)} + \underline{H}_{(2)} \qquad \text{(B-68)}$$

$$\underline{H}_{(1)} = \underline{H}_{1-0} + \underline{H}_{1-1} \qquad \text{(B-69)}$$

$$\underline{H}_{(3)} = \underline{H}_{3-0} + 3\underline{H}_{3-1} + 3\underline{H}_{3-2} + \underline{H}_{3-3} \qquad \text{(B-70)}$$

Then the optimum output f(t) is given by

$$f = \underline{H}[x]$$

$$= \left(\underline{H} * \left(\underline{L}_1^{-1} + \underline{N}^{-1} \right) \right) [y] \tag{B-71}$$

Working out the impulse responses of $\underline{H}$ by means of Eq. 231 shows that

$$\underline{H} * \underline{L}_1^{-1} * \underline{N}^{-1} = \underline{M} * \underline{K}_1 * \underline{N}^{-1} \tag{B-72}$$

The detailed nature of $\underline{M}$ and $\underline{K}_1$ are given in Eqs. 235-239 in example 8.

Acknowledgment

The writer is deeply appreciative of the continual guidance and assistance given to him by Professor Y. W. Lee. He would also like to acknowledge the support of the Department of Electrical Engineering and the Research Laboratory of Electronics, M. I. T., and many helpful discussions with several members of the Laboratory.

References

1. V. Volterra, Theory of Functionals and of Integral and Integro-Differential Equations (Blackie and Sons, Ltd., London, 1930).
2. N. Wiener, Response of a non-linear device to noise, Report 129, Radiation Laboratory, M.I.T., April 6, 1942.
3. A. G. Bose, A Theory of Nonlinear Systems, Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956.
4. M. B. Brilliant, Theory of the Analysis of Nonlinear Systems, Technical Report 345, Research Laboratory of Electronics, M.I.T., March 3, 1958.
5. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press, Cambridge, Massachusetts, and John Wiley and Sons, Inc., New York, 1958).
6. J. F. Barrett, The use of functionals in the analysis of non-linear physical systems, Statistical Advisory Unit Report 1/57, Ministry of Supply, Great Britain, 1957.
7. J. J. Stoker, Nonlinear Vibrations in Mechanical and Electrical Systems (Interscience Publishers, Inc., New York, 1950).
8. J. G. Truxal, Automatic Feedback Control System Synthesis (McGraw-Hill Book Company, Inc., New York, 1955).
9. M. F. Gardner and J. L. Barnes, Transients in Linear Systems, Vol. I (John Wiley and Sons, Inc., New York, 1942).
10. G. D. Zames, Nonlinear operators – cascading, inversion, and feedback, Quarterly Progress Report No. 53, Research Laboratory of Electronics, M.I.T., April 15, 1959, p. 93.
11. F. B. Hildebrand, Methods of Applied Mathematics (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1952).