#4

THEORY AND APPLICATION OF THE SEPARABLE CLASS OF RANDOM PROCESSES

ALBERT H. NUTTALL

TECHNICAL REPORT 343

MAY 26, 1958

Loan Copy Only

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-64637, Department of the Army Task 3-99-06-108 and Project 3-99-00-100.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 343 May 26, 1958

THEORY AND APPLICATION OF THE SEPARABLE CLASS OF RANDOM PROCESSES

Albert H. Nuttall

Submitted to the Department of Electrical Engineering, May 19, 1958 in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

The separable class of random processes is defined as that class of random processes for which the g-function,

$$g(x_2, \tau) = \int_{-\infty}^{\infty} (x_1 - \mu)\, p(x_1, x_2; \tau)\, dx_1$$

separates into the product of two functions, one a function only of x_2, the other a function only of τ. The second-order probability density function of the process is $p(x_1, x_2; \tau)$ and μ is its mean. Various methods of determining whether a random process is separable are developed, and basic properties of the separable class are derived.

It is proved that the separability of a random process that is passed through a nonlinear no-memory device is a necessary and sufficient condition for the input-output crosscovariance function to be proportional to the input autocovariance function, whatever nonlinear device is used. The uses of this invariance property are pointed out.

If a nonlinear no-memory device is replaced by a linear memory-capable network, so as to minimize the mean-square difference between the two outputs for the same separable input process, analysis shows that the optimum linear network has no memory. Simple relations among correlation functions for these circuits are also derived.

Some results on Markov processes and best estimate procedure are derived, important examples of separable processes are given, and possible generalizations of separability are stated.

Table of Contents

Table of Contents (continued)

GLOSSARY

Symbol	Definition
$p_x(x, t)$	First-order probability density function of x(t)
$p_x(x_1, t_1; \ldots ; x_n, t_n)$	n^{th}-order probability density function of x(t)
$p(x, t_1; y, t_2)$	Joint probability density function of x(t) and y(t)
$\mu_x(t)$	Mean value of x(t)
$\psi_x(t_1, t_2)$	Autocorrelation function of x(t)
$\phi_x(t_1, t_2)$	Autocovariance function of x(t)
$\sigma_x^2(t)$	Variance of x(t)
$\rho_x(t_1, t_2)$	Normalized covariance function (correlation coefficient) of x(t)
$f_x(\xi_1, t_1; \ldots ; \xi_n, t_n)$	n^{th}-order characteristic function of x(t)
$G_x(f)$	Frequency spectrum of stationary x(t)
Horizontal bar	Mean value

I. DEFINITION AND EQUIVALENT FORMULATIONS OF THE SEPARABLE CLASS OF RANDOM PROCESSES

1.1 INTRODUCTION

Many problems in electrical communication theory require statistical (probabilistic) methods of analysis in order to obtain useful measures of performance. For example, the prediction and filtering of some waveforms requires knowledge of correlation functions (1). Also, a quantitative measure of the rate of transmission of information between two points requires joint probability density functions (2). Although the statistical approach is the best one in many cases, the computations are often extremely involved and cannot be carried through unless a great deal of time and money is spent on the apparatus for computing, or special properties of certain cases are used. For instance, the useful properties of the Markov class of random processes have facilitated computations for some analysis problems.

In this report we define a property of partially integrated second-order statistics which we shall use to classify random processes into either separable or nonseparable classes. One of the purposes of this classification is to point out the simplicity of analysis of some communication networks when they are excited by the separable class of random processes.

The classification of random processes has been found to be a useful means of dealing with some electrical engineering analysis problems. In this connection, the Markov class of random processes has been found exceedingly useful. Thus, instead of working an analysis problem through in detail for a special random process, it may be that the problem can be worked through for a class of random processes. Indeed, the ability to work a problem, for a particular random process, may not be a virtue of that particular case, but of a broader class of random processes, all of which possess the same useful (and perhaps simple) properties. Such is found to be the case for the separable class of random processes in several engineering applications.

The establishment of the separable class of random processes may point the way to other classes of random processes which are useful for different analysis problems. In fact, in this report, we formulate some conditions, in addition to the separable classes' conditions, that are functions of higher-order statistics, and use them for classifying random processes for different purposes. The classification of a process does not depend on the use to which it is put. Rather, the uses of the class are investigated separately; the uses of the separable class are amply demonstrated here.

Knowledge of the fundamental rules of probability theory is necessary for understanding most of the present work and it will be assumed of the reader. The statistical quantities that are used are defined in the glossary.

The practice of presenting the simpler versions of a theory will be followed, for the most part, and generalizations will be made at the end of each section. For example, we present the analysis of stationary and time-invariant cases first, and then outline the general results that are applicable to nonstationary and time-variant cases afterwards.

1.2 FORMULATION OF SEPARABILITY

Let $p(x_1, x_2; \tau)$ be the second-order probability density function of a (stationary) random process. Define the g-function as

$$g(x_2, \tau) = \int (x_1 - \mu)\, p(x_1, x_2; \tau)\, dx_1 \qquad (1)$$

(All integrals without limits are over the range $(-\infty, \infty)$.) Notice that the g-function is determined by an integral on second-order statistics and involves no higher-order statistics. Now suppose that the g-function separates as

$$g(x_2, \tau) = g_1(x_2)\, g_2(\tau) \qquad (2)$$

for all x_2, τ. (We could now replace the variable x_2 by x but we retain the x_2 in order to keep symmetry of some formulas.) We then call the random process a <u>separable random process</u>. Now, since the autocovariance function of a stationary process $x(t)$ is

$$\begin{aligned}
\phi(\tau) &= \overline{[x(t) - \mu]\,[x(t+\tau) - \mu]} \\
&= \overline{[x(t) - \mu]\, x(t+\tau)} \\
&= \iint (x_1 - \mu) x_2\, p(x_1, x_2; \tau)\, dx_1\, dx_2 \\
&= \int x_2\, g(x_2, \tau)\, dx_2
\end{aligned}$$

where $p(x_1, x_2; \tau)$ is the second-order probability density function of the $x(t)$ process, we see that if Eq. 2 holds, we have

$$\phi(\tau) = g_2(\tau) \int x_2\, g_1(x_2)\, dx_2$$

Then

$$g_2(\tau) = \frac{\phi(0)}{\int x_2\, g_1(x_2)\, dx_2}\, \rho(\tau) = g_2(0)\, \rho(\tau)$$

where $\rho(\tau)$ is the normalized autocovariance function, or correlation coefficient, of the process. Thus, if the g-function separates as in Eq. 2, the function of τ must be a constant multiplied by $\rho(\tau)$. Now let us look at Eq. 1 for $\tau = 0$. Since, following Wang and Uhlenbeck (3),

$$p(x_1, x_2; 0) = p(x_1)\, \delta(x_2 - x_1)$$

we have

$$g(x_2, 0) = (x_2 - \mu)\, p(x_2) = g_1(x_2)\, g_2(0)$$

where $p(x_2)$ is the first-order probability density function of the process. Therefore

$$g_1(x_2) = \frac{(x_2 - \mu)p(x_2)}{g_2(0)}$$

and so

$$g(x_2, \tau) = (x_2 - \mu)p(x_2)\rho(\tau) \tag{3}$$

Incidentally, the same separation holds for the integral on $x_2 - \mu$:

$$\int (x_2 - \mu)p(x_1, x_2; \tau)\, dx_2 = \int (x_2 - \mu)p(x_2, x_1; -\tau)\, dx_2$$

$$= (x_1 - \mu)p(x_1)\rho(-\tau) = (x_1 - \mu)p(x_1)\rho(\tau)$$

Thus

> if a process is separable, the g-function must split up into the product of the correlation coefficient and a simple first-order statistic involving only the first-order probability density function of the process.

This is an extremely simple and useful property, as will become apparent. Hence the determination of whether a random process is separable or not can be easily determined by finding the second-order probability density function of the process, substituting in Eq. 1, and seeing if Eq. 3 holds. However, in some cases, the statistics relating to the process may be given in a different form, perhaps in a form involving the characteristic function. We shall derive the relation equivalent to Eq. 3 for the characteristic function of a separable process. The second-order characteristic function $f(\xi_1, \xi_2; \tau)$ of a random process x(t) is defined here as

$$f(\xi_1, \xi_2; \tau) = \overline{e^{j\xi_1[x(t)-\mu]}\, e^{j\xi_2[x(t+\tau)-\mu]}}$$

$$= \iint e^{j\xi_1(x_1-\mu)}\, e^{j\xi_2(x_2-\mu)}\, p(x_1, x_2; \tau)\, dx_1\, dx_2$$

This definition differs from the usual definition in the subtraction of the mean μ in the exponent. The justification for this definition lies in the simplicity of the formulas given in this report. Let us define a G-function as

$$G(\xi_2, \tau) = \left.\frac{\partial f(\xi_1, \xi_2; \tau)}{\partial \xi_1}\right|_{\xi_1=0}$$

$$= j \iint (x_1 - \mu)\, e^{j\xi_2(x_2-\mu)}\, p(x_1, x_2; \tau)\, dx_1\, dx_2$$

$$= j \int e^{j\xi_2(x_2-\mu)}\, g(x_2, \tau)\, dx_2 \tag{4}$$

This formula relates G and g for all processes regardless of separability. Since it is a Fourier transform, we can obtain g from G (the inverse of Eq. 4) as

$$g(x_2, \tau) = \frac{1}{j2\pi} \int e^{-j\xi_2(x_2-\mu)} G(\xi_2, \tau)\, d\xi_2 \tag{5}$$

It is now obvious from Eq. 4 that if g is separable, so also is G. Conversely, from Eq. 5, if G separates, so also does g. Either g and G both separate or neither does. Now G is found by a single partial derivative of the second-order characteristic function, whereas g is determined by a single integral of the second-order probability density function. Since derivatives are usually easier to effect, the determination of G might well be the point at which to ascertain the separability of the process. On the other hand, if $p(x_1, x_2; \tau)$ is more easily found than $f(\xi_1, \xi_2; \tau)$, determination of g might be easier. Each case must be investigated individually.

We have previously found that if g is separable, it is of the form given in Eq. 3. We substitute this relation in Eq. 4 to determine what form G must take for a separable process. We have

$$G(\xi_2, \tau) = j\rho(\tau) \int e^{j\xi_2(x_2-\mu)} (x_2 - \mu) p(x_2)\, dx_2$$

The first-order characteristic function $f(\xi_2)$ is defined here as

$$f(\xi_2) = \overline{e^{j\xi_2[x(t)-\mu]}}$$

$$= \int e^{j\xi_2(x_2-\mu)} p(x_2)\, dx_2$$

Therefore

$$f'(\xi_2) = j \int (x_2 - \mu)\, e^{j\xi_2(x_2-\mu)} p(x_2)\, dx_2$$

where the prime denotes a derivative with respect to the argument of f. We see therefore that G takes the form

$$G(\xi_2, \tau) = f'(\xi_2)\rho(\tau) \tag{6}$$

Then if a process is separable, the G-function must split up into the product of the derivative of the first-order characteristic function and the correlation coefficient.

This very simple general formula is found to be exceedingly useful for determining the basic properties of the separable class, as shown in Section II. Indeed, without this formula, the content of Section II could not have been worked out. Equation 6 is recognized as being the counterpart of Eq. 3 in a different domain.

Still another method of determining separability can be obtained. From Eq. 4 we have

$$G(\xi_2, \tau) = \overline{j\left[x(t)-\mu\right] e^{j\xi_2\left[x(t+\tau)-\mu\right]}}$$

$$= \overline{j\left[x(t)-\mu\right] \sum_{n=0}^{\infty} \frac{(j\xi_2)^n}{n!} x^n(t+\tau)\, e^{-j\xi_2\mu}}$$

$$= j\, e^{-j\xi_2\mu} \sum_{n=0}^{\infty} \frac{(j\xi_2)^n}{n!} \overline{\left[x(t)-\mu\right] x^n(t+\tau)}$$

Now if

$$\overline{\left[x(t)-\mu\right] x^n(t+\tau)} = b_n \rho(\tau) \qquad \text{for all } n \tag{7}$$

where b_n is a real number independent of τ, we have

$$G(\xi_2, \tau) = \rho(\tau) j \sum_{n=0}^{\infty} b_n \frac{(j\xi_2)^n}{n!} e^{-j\xi_2\mu}$$

which indicates a separable process. However, this line of attack requires that Eq. 7 be true, which is a sufficient but not a necessary condition. We may have a separable process for which Eq. 7 is not true merely because of the nonexistence of the left-hand side of Eq. 7 for large n. Thus separability and Eq. 7 are not identical; separability is a much more lenient condition. However, if the left-hand side of Eq. 7 exists for a certain n for a separable process, it must be the right-hand side of Eq. 7 for the same n. Therefore

$$\overline{\left[x(t)-\mu\right] x^n(t+\tau)} = \overline{x(t)x^n(t+\tau)} - \overline{x(t)}\,\overline{x^n(t+\tau)}$$

$$= \int x_2^n g(x_2, \tau)\, dx_2 = \rho(\tau) \int x_2^n (x_2 - \mu) p(x_2)\, dx_2$$

$$= b_n \rho(\tau)$$

Thus, if the waveform of x(t) is known, its substitution in the left-hand side of Eq. 7 can be made, the quantity determined for various n, compared with the right-hand side of Eq. 7, and separability ascertained. This test fails, as we have mentioned before, when and only when the left-hand side of Eq. 7 does not exist for some n. Since this approach does not require the explicit determination of the second-order probability density function or the characteristic function, it could, in some cases, be the simplest way of fixing separability.

In all that has gone before, the a-c component of the random process was used throughout. The reason for this will become apparent in Section III when we deal with covariance functions under distortion. The general theory could have been carried through with the mean present, of course, but future results dictate the present formulation of separability.

Notice that the requirement of separability of a process is a restriction not directly on the form of the second-order probability density function $p(x_1, x_2; \tau)$ but on an integral of it. This leaves open the possibility for a wide variety of separable random processes, and in Section VI we see that this is indeed so.

1.3 EXTENSION OF SEPARABILITY TO TWO RANDOM PROCESSES

We extend our definition of separability to two (stationary) random processes as follows: Let $p(x, y; \tau)$ be the joint probability density function for the two random processes x(t) and y(t). We define the g-function with respect to the y(t) process as

$$g(y, \tau) = \int (x - \mu_x) p(x, y; \tau)\, dx$$

If $g(y, \tau)$ separates as

$$g(y, \tau) = g_1(y) g_2(\tau)$$

we say that the x(t) process is separable with respect to the y(t) process. It then follows that, since

$$\phi_{xy}(\tau) = \int (y - \mu_y) g(y, \tau)\, dy$$

we have

$$g_2(\tau) = C\phi_{xy}(\tau)$$

where $\phi_{xy}(\tau)$ is the crosscovariance function of the processes x(t) and y(t), and C is a constant. However, in this case, we are unable to say anything specific about the form of $g_1(y)$. Whereas the second-order probability density function of a process has in it a delta function for zero shift, no such relation holds true for the joint probability density function of two random processes at any value of shift. Thus the results for two processes, although they are more general, are less specific.

For the G-function with respect to the y(t) process,

$$G(\xi_y, \tau) = \left. \frac{\partial f(\xi_x, \xi_y; \tau)}{\partial \xi_x} \right|_{\xi_x = 0}$$

we find that separability implies that

$$G(\xi_y, \tau) = \rho(\tau) G_1(\xi_y)$$

where the form of $G_1(\xi_y)$ is, in general, unknown. The same comments made above apply. Note that separability is easily determined, however, even though the general form of separation is not known. The fact that separability itself holds can be quite useful, even though the exact functional form, which depends on the particular case that is being investigated, is unknown.

Analogous to Eq. 7,

$$\overline{\left[x(t) - \mu_x\right] y^n(t+\tau)} = b_n \phi_{xy}(\tau)$$

for all n is a sufficient condition for the x(t) process to be separable with respect to the y(t) process. Again, this relation is not necessary; high-order averages may not exist.

It is obvious that separability can be ascertained with respect to either process, although one process may be separable with respect to the other without the converse holding true. Then, too, a process x(t) may be separable with respect to a process y(t) but not with respect to a process z(t). All these relations are functions of the exact statistics involved, and each must be investigated on its own merit.

1.4 EXTENSION TO NONSTATIONARY PROCESSES

For nonstationary (single) random processes, we define a g-function as

$$g(x_2; t_1, t_2) = \int \left[x_1 - \mu(t_1)\right] p(x_1, t_1; x_2, t_2)\, dx_1$$

and formulate separability as

$$g(x_2; t_1, t_2) = g_1(x_2, t_2) g_2(t_1, t_2)$$

Then by reasoning that is strictly analogous to that leading to Eq. 3, we find that

$$g(x_2; t_1, t_2) = \frac{\phi(t_1, t_2)}{\phi(t_2, t_2)} \left[x_2 - \mu(t_2)\right] p(x_2, t_2)$$

irrespective of the particular process used. Detailed computations with this g-function are no more difficult than with the stationary case, since the time instants t_1 and t_2 are mere dummy variables in any integration formulas for averages. Lack of stationarity adds no great complications to statistical problems with separable processes; nonstationary processes can be used as freely as stationary processes.

II. BASIC PROPERTIES OF THE SEPARABLE CLASS OF RANDOM PROCESSES

2.1 SUMS OF SEPARABLE PROCESSES

In communication networks the addition of two random processes occurs very frequently. In fact, since noise is present in every part of every physical piece of apparatus, the resultant quantity at the "output" must be the superposition of several noise voltages. In many cases, the addition of two processes (voltages) in the first stage, either intentional or otherwise, has a marked effect on the performance of the over-all system. Also, in communication across great distances, random atmospheric disturbances added to the transmitted signal constitute the received waveform that the receiver has to work on.

It is thus apparent that the study of the sum of two random processes is important for evaluating the performance of a communication system. Since separable processes will be found useful processes with which to excite networks by virtue of their simple analytic properties, it is of the utmost importance to determine under what conditions the sum of two (separable) processes is separable. Better still, we would like to find out what the necessary and sufficient conditions are in order that the sum of two separable random processes be separable.

Suppose we add two stationary processes x(t) and y(t) together to form a third stationary process z(t):

$$z(t) = x(t) + y(t)$$

In order to determine whether or not z(t) is a separable process, we have to determine

$$f_z(\xi_1, \xi_2;\tau) = \overline{e^{j\xi_1[z(t)-\mu_z]}\, e^{j\xi_2[z(t+\tau)-\mu_z]}}$$

$$= \iiiint e^{j\xi_1[x_1+y_1-\mu_x-\mu_y]}\, e^{j\xi_2[x_2+y_2-\mu_x-\mu_y]}\, p(x_1, y_1; x_2, y_2;\tau)\, dx_1\, dy_1\, dx_2\, dy_2$$

where $p(x_1, y_1; x_2, y_2;\tau)$ is the joint second-order probability density function of the two processes x(t) and y(t), and then take a derivative to determine $G_z(\xi_2, \tau)$. (In the rest of this section, subscripts will be put on the various statistical parameters in order to distinguish them.) Now in order to form any conclusions about the process z(t), we need further knowledge about the joint second-order probability density function. Instead of assuming specific forms for this probability density function, we assume that we are adding two independent processes, for which it is true that

$$p(x_1, y_1; x_2, y_2;\tau) = p_x(x_1, x_2;\tau)p_y(y_1, y_2;\tau)$$

where $p_x(x_1, x_2;\tau)$ and $p_y(y_1, y_2; \tau)$ are the second-order probability density functions for the x(t) and y(t) processes, respectively. Then

$$f_z(\xi_1, \xi_2; \tau) = f_x(\xi_1, \xi_2; \tau) f_y(\xi_1, \xi_2; \tau)$$

and

$$G_z(\xi_2, \tau) = \left. \frac{\partial f_z(\xi_1, \xi_2; \tau)}{\partial \xi_1} \right|_{\xi_1 = 0}$$

$$= G_x(\xi_2, \tau) f_y(\xi_2) + G_y(\xi_2, \tau) f_x(\xi_2) \tag{8}$$

Now if the x(t) and y(t) processes are not separable, it would be a coincidence if the z(t) process were separable, as is evident from Eq. 8. Accordingly, we restrict ourselves to the case in which the two processes x(t) and y(t) are both separable. Then using Eq. 6, we have

$$G_z(\xi_2, \tau) = f'_x(\xi_2) f_y(\xi_2) \rho_x(\tau) + f'_y(\xi_2) f_x(\xi_2) \rho_y(\tau) \tag{9}$$

Two sufficient conditions for the sum process z(t) to be separable are immediately evident:

$$\rho_x(\tau) = \rho_y(\tau)$$

and

$$f_x(\xi_2) = f_y(\xi_2)$$

are sufficient conditions for z(t) to be separable. (The latter condition requires, incidentally, that the two processes have the same variance. We shall remove this restriction presently.) However, let us now determine the necessary conditions for z(t) to be separable under the assumption that the two additive processes are independent and separable. We assume, then, with the use of Eq. 6, that

$$G_z(\xi_2, \tau) = f'_z(\xi_2) \rho_z(\tau) \tag{10}$$

Now

$$f'_z(\xi_2) = \left[f_x(\xi_2) f_y(\xi_2) \right]' = f'_x(\xi_2) f_y(\xi_2) + f_x(\xi_2) f'_y(\xi_2)$$

and

$$\rho_z(\tau) = \frac{\phi_x(\tau) + \phi_y(\tau)}{\phi_x(0) + \phi_y(0)} = a\rho_x(\tau) + b\rho_y(\tau)$$

where

$$a = \frac{\phi_x(0)}{\phi_x(0) + \phi_y(0)}$$

and

$$b = \frac{\phi_y(0)}{\phi_x(0) + \phi_y(0)}$$

Therefore, equating Eqs. 9 and 10, we require that

$$[f'_x(\xi_2)f_y(\xi_2) + f_x(\xi_2)f'_y(\xi_2)]\,[a\rho_x(\tau) + b\rho_y(\tau)] = f'_x(\xi_2)f_y(\xi_2)\rho_x(\tau) + f_x(\xi_2)f'_y(\xi_2)\rho_y(\tau)$$

Regrouping terms, we arrive at

$$[\rho_x(\tau) - \rho_y(\tau)]\,[bf'_x(\xi_2)f_y(\xi_2) - af_x(\xi_2)f'_y(\xi_2)] = 0$$

Therefore we must have either

$$\rho_x(\tau) = \rho_y(\tau) \tag{11a}$$

or

$$bf'_x(\xi_2)f_y(\xi_2) = af_x(\xi_2)f'_y(\xi_2)$$

The first condition has already been described as a sufficient condition. The second condition can be manipulated into a more tractable form:

$$\left[f_x(\xi_2)\right]^{1/\sigma_x^2} = \left[f_y(\xi_2)\right]^{1/\sigma_y^2} \tag{11b}$$

This condition is easily verified to be a sufficient condition by substitution in Eq. 9. Thus two sufficient conditions for the sum of two independant separable processes x(t) and y(t) to be separable are

$$\rho_x(\tau) = \rho_y(\tau)$$

and

$$[f_x(\xi)]^{1/\sigma_x^2} = [f_y(\xi)]^{1/\sigma_y^2}$$

Also, at least one of these conditions is necessary in order for the sum to be separable.

Hence, two independent separable processes, with identical spectra, added together form a new separable process with the same spectrum. On the other hand, if the spectra are not identical, the first-order characteristic functions must satisfy Eq. 11b, a rather restrictive condition. An immediate conclusion is that the sum of two processes is generally nonseparable. (We shall generalize our definition of separability in Section VII to alleviate this behavior.)

In attempting to extend the previous results to the sum of more than two independent separable processes, necessary conditions are not obvious. However, two sufficient conditions follow easily: Let

$$z(t) = \sum_{m=1}^{N} x_m(t)$$

Then

$$f_z(\xi_1, \xi_2; \tau) = \prod_{m=1}^{N} f_m(\xi_1, \xi_2; \tau)$$

and we have

$$G_z(\xi_2, \tau) = \left. \frac{\partial f_z(\xi_1, \xi_2; \tau)}{\partial \xi_1} \right|_{\xi_1=0}$$

$$= \sum_{m=1}^{N} f_1^{(\delta_{1m})}(\xi_2) \ldots f_N^{(\delta_{Nm})}(\xi_2) \rho_m(\tau)$$

since all processes are separable. The symbol δ_{jk} is the Kronecker delta which satisfies the condition

$$\delta_{jk} = \begin{Bmatrix} 0, & j \neq k \\ 1, & j = k \end{Bmatrix}$$

and

$$f_m^{(0)}(\xi_2) = f_m(\xi_2), \; f_m^{(1)}(\xi_2) = f_m'(\xi_2)$$

Now, it is obvious that if

$$\rho_1(\tau) = \rho_2(\tau) = \ldots = \rho_N(\tau) \equiv \rho(\tau) \tag{12a}$$

then

$$G_z(\xi_2, \tau) = f_z'(\xi_2) \rho(\tau)$$

and z(t) is a separable process. Thus the sum of N independent separable processes with identical spectra is separable. Alternatively, if

$$f_1(\xi_2)^{1/\sigma_1^2} = f_2(\xi_2)^{1/\sigma_2^2} = \ldots = f_N(\xi_2)^{1/\sigma_N^2} \tag{12b}$$

then

$$G_z(\xi_2, \tau) = f_1(\xi_2)^{(\sigma_2^2 + \ldots + \sigma_N^2)/\sigma_1^2} f_1'(\xi_2) \frac{1}{\sigma_1^2} \sum_{m=1}^{N} \sigma_m^2 \rho_m(\tau)$$

$$= \frac{d}{d\xi_2} \left\{ f_1(\xi_2)^{(\sigma_1^2 + \ldots + \sigma_N^2)/\sigma_1^2} \right\} \rho_z(\tau)$$

$$= f_z'(\xi_2) \rho_z(\tau)$$

and z(t) is a separable process. Now Eqs. 12a and 12b are two sufficient conditions for the sum process to be separable, but they have not been derived as being necessary. Indeed, the ability to prove necessity for the sum of two processes was a result of the fortunate grouping of terms in the expression just above Eq. 11a. No such grouping seems possible for more than two processes, and necessity has not been achieved.

2.2 PRODUCTS OF SEPARABLE PROCESSES

Another operation that commonly arises in communication networks is that of multiplication. For instance, with mixers, a predominant term in the output is the product of the two inputs. Accordingly, we shall investigate the product of two processes for separability. Let us form a stationary product process z(t) by multiplying two stationary separable processes x(t) and y(t):

$$z(t) = x(t)y(t)$$

Then

$$f_z(\xi_1, \xi_2;\tau) = \overline{e^{j\xi_1[z(t)-\mu_z]}\, e^{j\xi_2[z(t+\tau)-\mu_z]}}$$

$$= \iiiint e^{j\xi_1(x_1y_1-\mu_z)}\, e^{j\xi_2(x_2y_2-\mu_z)}\, p(x_1, y_1; x_2, y_2;\tau)\, dx_1\, dy_1\, dx_2\, dy_2$$

Again, we shall assume that the x(t) and y(t) processes are independent in order to obtain results about the z(t) process without assuming specific forms for the joint second-order probability density function. Temporarily, we shall also assume that $\mu_x = \mu_y = 0$. Then $\mu_z = 0$, and

$$f_z(\xi_1, \xi_2;\tau) = \iiiint e^{j\xi_1x_1y_1}\, e^{j\xi_2x_2y_2}\, p_x(x_1, x_2;\tau)p_y(y_1, y_2;\tau)\, dx_1\, dy_1\, dx_2\, dy_2$$

Therefore

$$G_z(\xi_2, \tau) = j\iiiint x_1y_1\, e^{j\xi_2x_2y_2}\, p_x(x_1, x_2;\tau)p_y(y_1, y_2;\tau)\, dx_1\, dy_1\, dx_2\, dy_2$$

$$= j\iint e^{j\xi_2x_2y_2}\, x_2p_x(x_2)\rho_x(\tau)y_2p_y(y_2)\rho_y(\tau)\, dx_2\, dy_2$$

in which we have used Eq. 3 for both the x(t) and y(t) processes. Then

$$G_z(\xi_2, \tau) = f'_z(\xi_2)\rho_z(\tau)$$

where

$$\rho_z(\tau) = \rho_x(\tau)\rho_y(\tau)$$

and

$$f_z(\xi_2) = \iint e^{j\xi_2xy}\, p_x(x)p_y(y)\, dx\, dy \tag{13}$$

because of the independence of the two processes. Thus the z(t) process is always separable; that is

> the product of two zero-mean independent stationary separable processes is always separable, irrespective of the particular statistics involved such as power density spectra, characteristic functions, and so on.

A further immediate conclusion is that the product of <u>any</u> number of zero-mean independent separable processes is always separable, again, irrespective of the particular statistics.

The determination of necessary conditions for the product of two independent separable processes to be separable is quite involved and is somehow related to the linear dependence or independence of $\rho_x(\tau)$, $\rho_y(\tau)$, and $\rho_x(\tau)\rho_y(\tau)$. In general, zero means are not necessary, as may be seen by considering the product of two independent (random) square waves alternating between A and -A, with symmetric or unsymmetric first-order probabiliby density functions. (Example 5 of Section VI demonstrates that a square wave is a separable process.) However, if the processes being multiplied do not have zero means, conditions for separability to hold true are involved. We shall derive an equation that the statistics of the individual processes must satisfy in order that the product process be separable, for the special case in which one of the processes, say x(t), has a zero mean. This is a special case of the more general case of two nonzero means, but it demonstrates the difficulties involved.

Let $\mu_x = 0$. Since the two processes are independent, $\mu_z = 0$, and

$$\rho_z(\tau) = \frac{\rho_x(\tau)}{1+\frac{\mu_y^2}{\sigma_y^2}}\left[\rho_y(\tau)+\frac{\mu_y^2}{\sigma_y^2}\right]$$

Multiplying $\rho_z(\tau)$ by $f_z'(\xi_2)$, with $f_z(\xi_2)$ given by Eq. 13, we must have

$$G_z(\xi_2,\tau) = j\iint xy\, e^{j\xi_2 xy} p_x(x)p_y(y)\, dx\, dy\, \frac{\rho_x(\tau)}{1+\frac{\mu_y^2}{\sigma_y^2}}\left[\rho_y(\tau)+\frac{\mu_y^2}{\sigma_y^2}\right] \tag{14}$$

for all ξ_2, τ, if the z(t) process is to be separable. Now, actually, $G_z(\xi_2,\tau)$ for $\mu_x = 0$ is given by

$$\begin{aligned} G_z(\xi_2,\tau) &= j\iiiint x_1y_1\, e^{j\xi_2x_2y_2}\, p_x(x_1,x_2;\tau)p_y(y_1,y_2;\tau)\, dx_1\, dx_2\, dy_1\, dy_2 \\ &= j\rho_x(\tau)\iint e^{j\xi_2x_2y_2}\, x_2p_x(x_2)p_y(y_2)[\mu_y+(y-\mu_y)\rho_y(\tau)]\, dx_2\, dy_2 \end{aligned} \tag{15}$$

where we have used Eq. 3 and recognized that $\mu_x = 0$. Thus, for separability of the z(t) process, we require equality of Eqs. 14 and 15:

$$j\rho_x(\tau)\iint e^{j\xi_2 xy}\, p_x(x)p_y(y)\left[\left\{x(y-\mu_y) - \frac{xy}{1+\dfrac{\mu_y^2}{\sigma_y^2}}\right\}\rho_y(\tau) + \left\{x\mu_y - \frac{xy}{1+\dfrac{\sigma_y^2}{\mu_y^2}}\right\}\right] dx\, dy = 0$$

for all ξ_2, τ. Or, if we rewrite, we have

$$\rho_x(\tau)\left[C_1(\xi_2)\rho_y(\tau) + C_2(\xi_2)\right] = 0 \qquad \text{for all } \xi_2,\ \tau \tag{16}$$

where

$$C_1(\xi_2) = j\iint e^{j\xi_2 xy}\, p_x(x)p_y(y)\left[x(y-\mu_y) - \frac{xy}{1+\dfrac{\mu_y^2}{\sigma_y^2}}\right] dx\, dy$$

and

$$C_2(\xi_2) = j\iint e^{j\xi_2 xy}\, p_x(x)p_y(y)x\left[\mu_y - \frac{y}{1+\dfrac{\sigma_y^2}{\mu_y^2}}\right] dx\, dy$$

Now let us hold ξ_2 fixed in Eq. 16; then, since $\rho_y(\tau)$ varies with τ, (otherwise y(t) would be dc, and therefore unimportant), we must have

$$C_1(\xi_2) = C_2(\xi_2) = 0$$

for all ξ_2. Writing $C_2(\xi_2) = 0$ out, we must have

$$0 = j\iint e^{j\xi_2 xy}\, p_x(x)p_y(y)x\left[\mu_y - \frac{y}{1+\dfrac{\sigma_y^2}{\mu_y^2}}\right] dx\, dy \tag{17}$$

for all ξ_2. $C_1(\xi_2) = 0$ yields nothing new, since the integrands in $C_1(\xi_2)$ and $C_2(\xi_2)$ are negatives of each other. Expanding Eq. 17, we have

$$0 = j\int_0^{\infty}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right]p_y(y)\int x e^{j\xi_2 xy}\, p_x(x)\, dx\, dy$$

$$+\, j\int_{-\infty}^{0}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right]p_y(y)\int x e^{j\xi_2 xy}\, p_x(x)\, dx\, dy =$$

$$= j\int_0^{\infty}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right] p_y(y)\int \frac{u}{y^2} e^{j\xi_2 u}\, p_x\left(\frac{u}{y}\right) du\, dy$$

$$-j\int_{-\infty}^{0}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right] p_y(y)\int \frac{u}{y^2} e^{j\xi_2 u}\, p_x\left(\frac{u}{y}\right) du\, dy$$

where the minus in the last equation is due to the fact that y is negative in the range of integration. Interchanging integrations, we have

$$0 = j\int u\, e^{j\xi_2 u}\left\{\int_0^{\infty}\frac{1}{y^2}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right] p_y(y)p_x\left(\frac{u}{y}\right) dy\right.$$

$$\left.-\int_0^{\infty}\frac{1}{y^2}\left[\mu_y - \frac{y}{1+\sigma_y^2/\mu_y^2}\right] p_y(y)p_x\left(\frac{u}{y}\right) dy\right\} du, \text{ all } \xi_2$$

or

$$\int_0^{\infty}\frac{1}{y^2}\left[\mu_y - \frac{y}{1+\frac{\sigma_y^2}{\mu_y^2}}\right] p_y(y)p_x\left(\frac{u}{y}\right) dy = \int_{-\infty}^{0}\frac{1}{y^2}\left[\mu_y - \frac{y}{1+\frac{\sigma_y^2}{\mu_y^2}}\right] p_y(y)p_x\left(\frac{u}{y}\right) dy$$

for almost all u. This condition is a rather stringent condition on the statistics but we have not been able to simplify it.

It may appear disturbing, at first, that the mean of a process should have so much to do with the separability of a product process. However, by looking first at the product process formed by two zero-mean processes, $x(t)$ and $y(t)$,

$$z_1(t) = x(t)y(t)$$

and then at the new product process obtained by adding a dc component to one of the processes, say the $y(t)$ process,

$$z_2(t) = x(t)\left[y(t)+\mu\right] = x(t)y(t)+\mu x(t)$$

we see that

$$z_2(t) = z_1(t) + \mu x(t)$$

Thus we have a sum of processes constituting $z_2(t)$. But we saw from our earlier results on sums of separable processes (and both $z_1(t)$ and $x(t)$ are separable), that, in general, the sum process is not separable. Granted that in the present case, $z_1(t)$ and $x(t)$ are not independent, yet the essential points are the same. Thus it is not too surprising that nonzero means can destroy the separability of the product process.

2.3 NONSTATIONARY PROCESSES

We now generalize to the case of nonstationary time series. By reasoning analogous to that for the time stationary case, we find that

$$\frac{\phi_x(t_1, t_2)}{\phi_x(t_2, t_2)} = \frac{\phi_y(t_1, t_2)}{\phi_y(t_2, t_2)}$$

or

$$\left[f_x(\xi, t)\right]^{1/\sigma_x^2(t)} = \left[f_y(\xi, t)\right]^{1/\sigma_y^2(t)}$$

are sufficient conditions in order that the sum process $z(t) = x(t) + y(t)$ be separable, if $x(t)$ and $y(t)$ are independent separable processes. Also, at least one of these conditions is necessary in order that the sum process be separable.

For products of processes, we find that the product process $z(t) = x(t)y(t)$ is always separable if $x(t)$ and $y(t)$ are independent, separable, and have zero means. Thus

$$\mu_x(t) = \mu_y(t) = 0$$

Again, necessary conditions seem to be very difficult to obtain.

In summary, whereas multiplying processes generally preserves separability, adding processes seldom does. This is unfortunate, because addition is a far more important operation in communication networks than multiplication.

III. INVARIANCE OF COVARIANCE FUNCTIONS UNDER NONLINEAR TRANSFORMATIONS

3.1 THE INVARIANCE PROPERTY

In the transmission of information between two distant communicators, the detection of a signal in additive noise is all important. Various schemes have been devised for the reliable determination of whether or not a signal is present in a received waveform (4), and which of several possible signals is present (5). Many of these schemes depend on the determination of a crosscorrelation function, or equivalently, a crosscovariance function.

Since the determination of correlation functions is so important, we shall investigate in this section some of the useful properties of the separable class of random processes for correlation measurements, and point out potential ways of utilizing the separable class for simplified correlation and detection schemes. We shall also relate the present results to past work on the invariance of correlation functions under transformations.

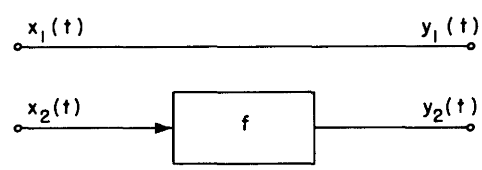

Fig. 1 Nonlinear no-memory system.

Consider the system of Fig. 1, in which the two stationary random processes $x_1(t)$ and $x_2(t)$ are used as inputs, and $y_1(t)$ and $y_2(t)$ are outputs. One of the inputs, $x_1(t)$, is passed through an all-pass network, and the other input, $x_2(t)$, is passed through a nonlinear no-memory network (which will be called a device) characterized by the function f. Hence

$$y_2(t) = f\left[x_2(t)\right]$$

The output $y_2(t)$ of the device f at time t thus depends solely on the input to the device at the same time instant t. Let us now define the input crosscovariance function as

$$\begin{aligned}\phi(\tau) &= \overline{\left[x_1(t) - \mu_1\right]\left[x_2(t+\tau) - \mu_2\right]} \\ &= \overline{\left[x_1(t) - \mu_1\right]x_2(t+\tau)} \\ &= \iint (x_1 - \mu_1)x_2 p(x_1, x_2;\tau)\, dx_1\, dx_2 \end{aligned} \tag{18}$$

where $p(x_1, x_2; \tau)$ is the joint probability density function of the inputs $x_1(t)$ and $x_2(t)$, and define the output crosscovariance function as

$$\Phi_f(\tau) = \overline{\left[y_1(t) - \overline{y_1(t)}\right]\left[y_2(t+\tau) - \overline{y_2(t+\tau)}\right]}$$

$$= \overline{\left[y_1(t) - \overline{y_1(t)}\right] y_2(t+\tau)} = \overline{\left[x_1(t) - \mu_1\right] f\left[x_2(t+\tau)\right]}$$

$$= \iint (x_1 - \mu_1) f(x_2) p(x_1, x_2; \tau)\, dx_1\, dx_2 \tag{19}$$

A subscript is put on $\Phi(\tau)$ in order to indicate the dependence of the crosscovariance function on the particular device f that is used. Now, in general, there is no relation between $\phi(\tau)$ and $\Phi_f(\tau)$. However, for the case in which the $x_1(t)$ process is separable with respect to the $x_2(t)$ process,

$$g(x_2, \tau) = g_1(x_2)\phi(\tau)$$

we see from Eq. 19 that

$$\Phi_f(\tau) = \phi(\tau) \int f(x_2) g_1(x_2)\, dx_2$$

$$= C_f \phi(\tau) \qquad \text{for all } f \text{ and } \tau \tag{20}$$

where C_f is a number depending on the nonlinear device and the input statistics. Thus,

> regardless of the nonlinear device used, the input and output crosscovariance functions are identical, except for a scale factor, when the undistorted process is separable with respect to the "distorted" input process.

The scale factor depends on the nonlinear device and the input statistics. However, for a fixed pair of inputs and a fixed nonlinear device, this scale factor does not change with τ.

This behavior of the covariance functions is called the invariance property. Note that we include, as a special case, $x_1(t) = x_2(t)$. In this case, the invariance property relates the input-output crosscovariance function to the input autocovariance function if the input is a separable process. In the case of one input, also, the simpler results for the separable class, as derived in Section I are applicable. That is, since $x_1(t) = x_2(t) \equiv x(t)$,

$$g(x_2, \tau) = (x_2 - \mu) p(x_2) \rho(\tau)$$

where $p(x_2)$ is the first-order probability density function of the input process $x(t)$, and

$$C_f = \frac{1}{\sigma^2} \int f(x_2)(x_2 - \mu) p(x_2)\, dx_2 \tag{21}$$

No such simple formula as Eq. 21 holds when $x_1(t) \neq x_2(t)$. In this latter case, $g(x_2, \tau)$ must be found from (1).

3.2 USE OF THE INVARIANCE PROPERTY

The importance of the invariance property can be seen in several ways. Suppose, for instance, that we wish to measure the crosscovariance function of two random processes $x_1(t)$ and $x_2(t)$. We find, in amplifying these processes that one of the processes undergoes some unintentional nonlinear no-memory distortion. If the undistorted input process is separable with respect to the "distorted" input process, the nonlinearity gives us no trouble whatsoever, and we can measure the crosscovariance function after amplification (at the output) and be assured that what we have, in fact, computed is what we set out to get.

Or alternatively, we may deliberately put in a nonlinear no-memory device to operate on, say, $x_2(t)$ to simplify covariance measurements. For example, if f were the perfect peak clipper,

$$f(x) = \begin{Bmatrix} -1, & x < 0 \\ +1, & x \geqslant 0 \end{Bmatrix}$$

then the computation of the output crosscovariance function is effected by multiplying a waveform $y_1(t)$ by a second delayed waveform, $y_2(t+\tau)$, which takes on only two values, ± 1, and by averaging. This follows by virtue of Eq. 20. But in this case multiplication can be replaced by gating. The output $y_1(t)$ of a network is controlled by the polarity of the waveform $y_2(t+\tau)$. Thus we have a simplified correlator that replaces the multiplier that is used in the conventional correlator by a gating network. In both schemes, a delay line must be present together with an averager, and so nothing is gained along this line. However, the multiplier is the hardest component to construct and we can actually eliminate this component in favor of a gating network that is easier to construct. This applies only for separable processes.

3.3 DISCUSSION AND INTERPRETATION OF THE INVARIANCE PROPERTY

In leading up to Eq. 20 we used the fact that $x_1(t)$ was separable with respect to $x_2(t)$. Thus separability was a sufficient condition for the invariance property to hold. The sufficiency of certain classes of processes for yielding the invariance property has been investigated by several people. Bussgang (6) showed that the Gaussian process satisfied the invariance property, and Luce (7) generalized to the class of separable processes. (Luce could not prove the necessity of the separable class for the invariance property to hold; in fact, he tried to generalize the Gaussian result to a wider class. However, one of the consistency relations that must be satisfied by a joint probability density function is violated in his work, and hence the proposed class is invalidated.) Barrett and Lampard (8) then generalized to a class of processes that could be expanded in a special single series, and Brown (9) later generalized to a special double series. We shall have more to say about these classes later. In any event, all previous results have been directed toward sufficiency classes. The question obviously arises about what is the most general class of processes for the invariance property to hold. In section 3.4 we prove rigorously that a necessary and sufficient condition for the invariance property to hold is that the g-function separate. Then, since separability is both necessary and sufficient, we know that determination of the g-function (or G-function) is the only necessary calculation for checking whether or not the invariance property holds. This eliminates the degree of coincidence which has accompanied the invariance property thus far, and states in definite terms what must be true of the input processes.

Since the proof of necessity in section 3.4 is lengthy, we present here an alternative approach to the problem which, as well as being shorter, gives more insight into the problem, and points out experimental limitations in determining correlation functions by use of separability. Referring to Fig. 1, let us assume nothing about separability of the inputs, but, rather, let us suppose that the output crosscovariance function $\Phi_f(\tau)$ is proportional to the input crosscovariance function $\phi(\tau)$ — not for all nonlinear devices f but only for a particular device f_1. That is,

$$\Phi_{f_1}(\tau) = \int f_1(x_2) g(x_2, \tau)\, dx_2 = C_{f_1}\phi(\tau)$$

for all τ, but nothing is known (at the moment) about $\Phi_{f_2}(\tau)$ where $f_2 \neq f_1$. Then

$$C_{f_1} = \frac{\Phi_{f_1}(\tau_1)}{\phi(\tau_1)} = \frac{1}{\phi(\tau_1)}\int f_1(x_2) g(x_2, \tau_1)\, dx_2$$

where τ_1 is any τ with the property that $\phi(\tau_1) \neq 0$. Therefore

$$\int f_1(x_2)\left[g(x_2, \tau) - \frac{\phi(\tau)}{\phi(\tau_1)} g(x_2, \tau_1)\right] dx_2 = 0$$

for all τ. Thus, since τ_1 is some fixed number, if we let

$$g_1(x_2) = \frac{g(x_2, \tau_1)}{\phi(\tau_1)}$$

and

$$v(x_2, \tau) = g(x_2, \tau) - g_1(x_2)\phi(\tau) \tag{22}$$

we have

$$\int f_1(x_2) v(x_2, \tau)\, dx_2 = 0$$

for all τ. That is, $v(x_2, \tau)$, which is determined solely by input statistics, is orthogonal, as a function of x_2, to the particular function f_1, for all τ. Therefore satisfaction of the invariance property for a particular function f_1 implies very little about the function $v(x_2, \tau)$ — only orthogonality with f_1.

Let us now investigate the meaning of these results in terms of laboratory measurements. Suppose we find experimentally that the output crosscovariance function for a particular nonlinear distortion f_1 is proportional to the input crosscovariance function.

Then we have found out nothing about the separability of the input process but merely that $v(x_2, \tau)$, as defined by Eq. 22, is orthogonal to the particular nonlinear device f_1.

It is readily seen, however, that $v(x_2, \tau)$ can be tied down quite effectively if we specify that the covariance functions be proportional for a certain class C of nonlinear devices rather than for one particular nonlinear device. Then, for the class C, $v(x_2, \tau)$ is orthogonal to each and every one of the functions (or devices) in C. Now if C is a closed (or complete) system of functions (10), then $v(x_2, \tau)$ must be zero except for a set (in x_2) of zero measure, for any τ. If C is not a closed system, then nothing specific can be said about $v(x_2, \tau)$ except that it is orthogonal to all the functions in C; this, in the general case, allows $v(x_2, \tau)$ to take on widely different characteristics.

Interpreting, again, in terms of laboratory measurements, we are led by the results to the conclusion that since we can make only a finite number of measurements, we can never hope to prove experimentally that the inputs are separable. We can only state that $v(x_2, \tau)$ for that pair of inputs is orthogonal to all the nonlinear devices considered. Then in order to state separability of processes, we need to know something more about the input processes, such as how they were formed or the nature of the components producing them. Thus we see that the results of Section II are of importance in determining whether a process (or processes) is separable. Also, the various methods of determining separability, as given in Section I should find some use in this investigation.

We can now state that:

> if the invariance property holds for any closed system of functions, then the g-function is separable (almost everywhere).

The converse cannot be stated unequivocally; the separability of g is not enough. It may be that for a particular function in the closed system that is under consideration, the output covariance function does not exist for all τ. In such a case the invariance property cannot possibly hold. But let us define the class C_p of nonlinear devices as that class for which the output crosscovariance function exists for all τ. Then we alter the previous statement to read:

> The separability of the g-function is a necessary and sufficient condition for the invariance property to hold for any closed system of functions that is a subset of C_p.

By this approach we can derive both necessity and sufficiency. In section 3.4, we derive this relation in a manner that does not require the statement of a closed system of functions and is, therefore, more useful.

3.4 PROOF OF THE INVARIANCE PROPERTY

Consider the system of Fig. 1 with stationary input processes $x_1(t)$ and $x_2(t)$, and a time-invariant nonlinear no-memory network f. The output $y_2(t)$ is determined as

$$y_2(t) = f\left[x_2(t)\right]$$

The input crosscovariance function, defined as

$$\phi(\tau) = \overline{\left[x_1(t) - \mu_1\right]\left[x_2(t+\tau) - \mu_2\right]} = \overline{\left[x_1(t) - \mu_1\right]x_2(t+\tau)}$$
$$= \iint (x_1 - \mu_1)x_2 p(x_1, x_2; \tau)\, dx_1\, dx_2$$

is assumed to exist as a finite-valued Lebesgue double integral for all τ. That is, $(x_1 - \mu_1)x_2 p(x_1, x_2; \tau)$ is Lebesgue integrable for all τ. The input joint probability density function is $p(x_1, x_2; \tau)$.

The output crosscovariance function, defined as

$$\Phi_f(\tau) = \overline{\left[y_1(t) - \overline{y_1(t)}\right]\left[y_2(t+\tau) - \overline{y_2(t+\tau)}\right]} = \overline{\left[y_1(t) - \overline{y_1(t)}\right]y_2(t+\tau)}$$
$$= \overline{\left[x_1(t) - \mu_1\right]f\left[x_2(t+\tau)\right]}$$
$$= \iint (x_1 - \mu_1)f(x_2)p(x_1, x_2; \tau)\, dx_1\, dx_2 \tag{23}$$

exists as a Lebesgue double integral, for all τ, only for a class C_p of nonlinear devices f. That is, f is in C_p if and only if Eq. 23 exists for all τ. The class C_p depends only on the second-order probability density function $p(x_1, x_2; \tau)$. The subscript f on $\Phi_f(\tau)$ indicates the dependence of the output crosscovariance function on the particular nonlinear device that is under investigation.

Let us rewrite Eq. 23 as

$$\Phi_f(\tau) = \int f(x_2)g(x_2, \tau)\, dx_2 \tag{24}$$

where

$$g(x_2, \tau) = \int (x_1 - \mu_1)p(x_1, x_2; \tau)\, dx_1$$

We shall now prove the following theorem.

THEOREM:

$$\left.\begin{array}{l} g(x_2, \tau) = g_1(x_2)\phi(\tau) \\ \text{almost everywhere} \\ \text{in } x_2, \text{ for all } \tau \end{array}\right\} \begin{array}{c} \text{implies and} \\ \text{is implied} \\ \text{by} \end{array} \left\{\begin{array}{l} \Phi_f(\tau) = C_f\phi(\tau) \\ \text{for all f in } C_p, \\ \text{for all } \tau. \end{array}\right. \tag{25}$$

C_f is a constant dependent only on the particular nonlinear device f. The satisfaction of the left-hand side of Eqs. 25 will be called separability (of the g-function). The satisfaction of the right-hand side of Eqs. 25 will be called the invariance property.

Thus, the separability of the g-function is a necessary and sufficient condition for the invariance property to hold.

PROOF:

Sufficiency:

From Eq. 24, for any $f \in C_p$,

$$\Phi_f(\tau) = \int f(x_2)g(x_2, \tau)\, dx_2$$

$$= \int f(x_2)g_1(x_2)\phi(\tau)\, dx_2 = C_f\phi(\tau)$$

where

$$C_f = \int f(x_2)g_1(x_2)\, dx_2$$

C_f certainly exists by virtue of $f \epsilon C_p$; that is, Eq. 24 exists for all τ.

Necessity:

We have, by assumption, $\Phi_f(\tau) = C_f\phi(\tau)$, for any $f \epsilon C_p$, for all τ.

Suppose, now, that $\phi(\tau_1) \neq 0$, for some τ_1. (If this is not true, the following derivation needs few revisions to establish the separability of g.) Then

$$C_f = \frac{\Phi_f(\tau_1)}{\phi(\tau_1)}$$

for any $f \epsilon C_p$. Therefore

$$\Phi_f(\tau) = \Phi_f(\tau_1)\frac{\phi(\tau)}{\phi(\tau_1)}$$

for any $f \epsilon C_p$, for all τ. Therefore, from Eq. 24,

$$\int f(x_2)g(x_2, \tau)\, dx_2 = \frac{\phi(\tau)}{\phi(\tau_1)}\int f(x_2)g(x_2, \tau_1)\, dx_2$$

for any $f \epsilon C_p$, for all τ. Or, equivalently,

$$\int f(x_2)\left[g(x_2, \tau) - \frac{\phi(\tau)}{\phi(\tau_1)}g(x_2, \tau_1)\right]dx_2 = 0$$

for any $f \epsilon C_p$, for all τ.

Consider <u>any</u> fixed τ; call it τ_2. Then

$$\int f(x_2)\left[g(x_2, \tau_2) - \frac{\phi(\tau_2)}{\phi(\tau_1)}g(x_2, \tau_1)\right]dx_2 = 0$$

for any $f \epsilon C_p$. Let

$$v_{\tau_1\tau_2}(x_2) = g(x_2, \tau_2) - \frac{\phi(\tau_2)}{\phi(\tau_1)}g(x_2, \tau_1) \tag{26}$$

Then

$$\int f(x_2)v_{\tau_1\tau_2}(x_2)\, dx_2 = 0 \tag{27}$$

for any $f \epsilon C_p$. Choose

$$f(x_2) = \begin{cases} +|x_2| & \text{if } v_{\tau_1\tau_2}(x_2) > 0 \\ 0 & \text{if } v_{\tau_1\tau_2}(x_2) = 0 \\ -|x_2| & \text{if } v_{\tau_1\tau_2}(x_2) < 0 \end{cases} \tag{28}$$

This particular f belongs to C_p, since it is measurable because g is measurable. Also

$$|f(x_2)| \leq |x_2|$$

and therefore

$$|f(x_2)g(x_2, \tau)| \leq |x_2 g(x_2, \tau)|$$

for all τ. Now $|x_2 g(x_2, \tau)|$ is integrable for all τ, since $x_2 g(x_2, \tau)$ is integrable for all τ. Therefore $f(x_2)g(x_2, \tau)$ is integrable for all τ. Thus $f \in C_p$.

Thus with the use of f, as defined in Eq. 28, in Eq. 27, we have

$$\int \left| x_2 v_{\tau_1 \tau_2}(x_2) \right| dx_2 = 0$$

With the use of the lemma stated below, it follows that

$$x_2 v_{\tau_1 \tau_2}(x_2) = 0 \text{ almost everywhere}$$

Therefore

$$v_{\tau_1 \tau_2}(x_2) = 0 \text{ almost everywhere}$$

Recalling Eq. 26, we have

$$g(x_2, \tau_2) = \frac{\phi(\tau_2)}{\phi(\tau_1)} g(x_2, \tau_1) \text{ almost everywhere in } x_2$$

Since, now, τ_2 is <u>any</u> τ, we have

$$g(x_2, \tau) = \frac{g(x_2, \tau_1)}{\phi(\tau_1)} \phi(\tau) \text{ almost everywhere in } x_2, \text{ for all } \tau$$

And since τ_1 is a fixed number, $\frac{g(x_2, \tau_1)}{\phi(\tau_1)}$ depends only on x_2. We shall denote it by

$$g_1(x_2) = \frac{g(x_2, \tau_1)}{\phi(\tau_1)}$$

Then

$$g(x_2, \tau) = g_1(x_2)\phi(\tau) \text{ almost everywhere in } x_2, \text{ for all } \tau.$$

It is obvious from this proof that, instead of Eqs. 25, we can state:

$g(x_2, \tau) = g_1(x_2)\phi(\tau)$ almost everywhere in x_2, for $\tau \in S$ where S is an arbitrary set of the real line.	implies and is implied by	$\Phi_f(\tau) = C_f \phi(\tau)$ for all $f \in C_p$, for all $\tau \in S$

LEMMA:

Given:

$$\int_E f d\mu = 0 \qquad \text{with} \qquad f \geq 0 \text{ on } E$$

Let E_n be the subset of E on which $f > \frac{1}{n}$. $E_1 \subset E_2 \subset \ldots \subset E$. Then $\mu(E_n) = 0$ for all n. Otherwise, $\int_E f d\mu > 0$. Let

$$E'_1 = E_1, E'_2 = E_2 - E_1, E'_3 = E_3 - E_2, \ldots$$

Therefore

$$E'_m \cap E'_n = \phi \text{ if } m \neq n$$

where ϕ is the empty set. Then let

$$A = \bigcup_{n=1}^{\infty} E'_n$$

and

$$\mu(A) = \sum_{n=1}^{\infty} \mu(E'_n)$$

by the complete additivity of μ. Now

$$0 \leq \mu(E'_n) \leq \mu(E_n) = 0$$

Therefore $\mu(E'_n) = 0$, for all n, and $\mu(A) = 0$. However, A is the subset of E on which $f \neq 0$. Therefore $f = 0$ almost everywhere on E. If we let μ be Lebesgue measure, we have our particular case.

3.5 NECESSITY OF COVARIANCE FUNCTIONS

In Section I we defined the g-function for a single stationary process as an integral on the second-order probability density function of the input process with a factor $(x_1 - \mu)$:

$$g(x_2, \tau) = \int (x_1 - \mu) p(x_1, x_2; \tau)\, dx_1 \qquad (1)$$

The reason for this definition will now be explained with reference to two stationary input processes. One input is, as mentioned before, a special case. Consider, again, the system of Fig. 1. Suppose we define the input and output crosscorrelation functions, respectively, as

$$\psi(\tau) = \overline{x_1(t) x_2(t+\tau)}$$

and

$$\Psi_f(\tau) = \overline{y_1(t)y_2(t+\tau)}$$

$$= \overline{x_1(t)f\left[x_2(t+\tau)\right]}$$

Now if we require

$$\Psi_f(\tau) = C_f\psi(\tau) \tag{29}$$

for all τ, where C_f is a constant dependent only on f, then for f a constant, say K, we shall have

$$\mu_1 K = C_f\psi(\tau)$$

for all τ. This requires $\psi(\tau)$ to be a constant and is obviously much too restrictive. We therefore define the ac input and output crosscorrelation functions, respectively, as

$$\phi(\tau) = \overline{\left[x_1(t) - \mu_1\right]\left[x_2(t+\tau) - \mu_2\right]} = \psi(\tau) - \mu_1\mu_2$$

and

$$\Phi_f(\tau) = \overline{\left[y_1(t) - \overline{y_1(t)}\right]\left[y_2(t+\tau) - \overline{y_2(t+\tau)}\right]} = \Psi_f(\tau) - \mu_1\overline{y_2(t)}$$

These functions are also known as the crosscovariance functions, as explained in Appendix 1. The invariance property is now stated as

$$\Phi_f(\tau) = C_f\phi(\tau) \tag{30}$$

for all τ. This form is identical with the previous form, Eq. 29, when $\mu_1 = 0$, but does not suffer from the case of having f a constant. For we have

$$0 = C_f\phi(\tau)$$

and consequently $C_f = 0$. This leaves $\phi(\tau)$ completely arbitrary.

Then with $g(x_2, \tau)$ as defined in Eq. 1, with μ replaced by μ_1 (for two processes), we have

$$\phi(\tau) = \int x_2 g(x_2, \tau)\, dx_2$$

and

$$\Phi_f(\tau) = \int f(x_2) g(x_2, \tau)\, dx_2$$

and if $\Phi_f(\tau)$ and $\phi(\tau)$ are proportional, it is obvious that any relation that must hold true will involve $g(x_2, \tau)$, and not the joint probability density function $p(x_1, x_2; \tau)$ explicitly.

3.6 EXTENSION TO NONSTATIONARY PROCESSES

We now discuss the results that relate to time-variant statistics and a time-variant device f in Fig. 1. Now

$$y_2(t) = f\left[x_2(t), t\right]$$

which means that the nonlinear function relating $x_2(t)$ to $y_2(t)$ depends also upon the time t. For time-variant statistics we define the input crosscovariance function as

$$\phi(t_1, t_2) = \overline{\left[x_1(t_1) - \mu_1(t_1)\right]\left[x(t_2) - \mu_2(t_2)\right]}$$
$$= \int x_2 g(x_2; t_1, t_2)\, dx_2$$

where $p(x_1, t_1; x_2, t_2)$ is the joint input probability density function, and

$$g(x_2; t_1, t_2) = \int \left[x_1 - \mu_1(t_1)\right] p(x_1, t_1; x_2, t_2)\, dx_1$$

Also, the output crosscovariance function is

$$\Phi_f(t_1, t_2) = \overline{\left[y_1(t_1) - \overline{y_1(t_1)}\right]\left[y_2(t_2) - \overline{y_2(t_2)}\right]}$$
$$= \int f(x_2, t_2) g(x_2; t_1, t_2)\, dx_2$$

Now if $x_1(t)$ is separable with respect to $x_2(t)$, then

$$g(x_2; t_1, t_2) = g_1(x_2, t_2)\phi(t_1, t_2) \tag{31}$$

and

$$\Phi_f(t_1, t_2) = \phi(t_1, t_2)\int f(x_2, t_2) g_1(x_2, t_2)\, dx_2$$
$$= C_f(t_2)\phi(t_1, t_2) \tag{32}$$

Thus the covariance functions are proportional for fixed t_2. This is the invariance property for the time-variant case. We must include a time dependence in the scale factor C_f to allow either for a time-variant device or for input statistics that change with time. Conversely, if Eq. 32 is true for all t_1, t_2 and all f for which the left-hand side of Eq. 32 exists, we can show in a manner that is strictly analogous to that in section 3.4 that the g-function separates as in Eq. 31. Thus,

> for nonstationary processes and time-variant devices, separability of the g-function is a necessary and sufficient condition for the invariance property to hold.

If the input processes are the same, and separable, then from Section I,

$$g(x_2; t_1, t_2) = \phi(t_1, t_2)\frac{x_2 - \mu(t_2)}{\sigma^2(t_2)}\, p(x_2, t_2) \tag{33}$$

and the scale factor relating covariance functions becomes

$$C_f(t_2) = \int f(x_2, t_2)\frac{x_2 - \mu(t_2)}{\sigma^2(t_2)}\, p(x_2, t_2)\, dx_2 \tag{34}$$

These relations, Eqs. 33 and 34, readily reduce to those stated earlier for the time stationary case.

As a special case of Eq. 32, consider

$$f_n(x_2, t_2) = x_2^n$$

and let $\Phi_{f_n}(t_1, t_2)$ exist. Then it follows that

$$\overline{[x_1(t_1) - \mu_1(t_1)]x_2^n(t_2)} = C_{f_n}(t_2)\phi(t_1, t_2)$$

or

$$\overline{x_1(t_1)x_2^n(t_2)} - \overline{x_1(t_1)}\,\overline{x_2^n(t_2)} = C_{f_n}(t_2)\phi(t_1, t_2)$$

for all n for which the left-hand side exists. This is a convenient way of stating the invariance property, in that moments are readily interpretable.

One more variation is obvious with the invariance property: If we restrict ourselves to time-invariant networks but allow nonstationary inputs, we obtain

$$g(x_2; t_1, t_2) = g_1(x_2)\phi(t_1, t_2)$$

as a necessary and sufficient condition for the invariance property to hold. This relation is somewhat more restrictive than Eq. 31.

3.7 APPLICATION TO RADAR

We have already pointed out one important use of the invariance property and we now elaborate slightly on the method. Let us consider that we are transmitting a random signal s(t) and measuring the range of a target by correlating the return signal with a delayed version of the transmitted signal (5). Then our received signal will be

$$y(t) = As(t - \tau_1) + n(t) \tag{35}$$

where n(t) is additive independent noise (intentional or otherwise), A is a gain factor dependent on the range of the target, and τ_1, is the two-way delay to the target. [We assume there are no multiple reflections and that A does not vary over the time of transmission of s(t).] We now correlate y(t) with a delayed version of the stored signal, $s(t - \tau_2)$, where τ_2 is variable and at our control. Then the average output of the correlator is

$$\overline{y(t)s(t - \tau_2)} = \overline{[As(t - \tau_1) + n(t)]s(t - \tau_2)}$$

$$= \bar{A}\,\phi_s(\tau_1 - \tau_2)$$

in which we have let $\overline{s(t)} = \overline{n(t)} = 0$, which is the usual case. Now in some applications, the spectrum of the transmitted signal is deliberately shaped, so that the rate of decay of the correlation function indicates how far off our local estimate τ_2 of the actual delay τ_1 is, and also spurious indications of the delay are avoided. It is therefore all important that the output of the correlator yield something proportional to $\phi_s(\tau_1 - \tau_2)$. Now let us consider that before we crosscorrelate y(t) with $s(t - \tau_2)$, we deliberately distort the stored signal $s(t - \tau_2)$ by means of a nonlinear device f. Then, as the average output of our correlator, we obtain

$$\overline{y(t)f\left[s(t - \tau_2)\right]} = \overline{\left[As(t - \tau_1) + n(t)\right]f\left[s(t - \tau_2)\right]}$$

$$= \bar{A}\overline{s(t - \tau_1)f\left[s(t - \tau_2)\right]}$$

$$= \bar{A}\iint x_1 f(x_2) p_s(x_1, x_2; \tau_1 - \tau_2)\, dx_1\, dx_2$$

Now suppose that we use a separable process for the transmitted signal. Then the average output, with the use of Eqs. 1 and 3, is

$$\overline{y(t)f\left[s(t - \tau_2)\right]} = \bar{A}\rho_s(\tau_1 - \tau_2)\int f(x_2) x_2 p_s(x_2)\, dx_2$$

and we have preserved exactly the form of the correlation function, as we desired. In this arrangement we transmit a random signal s(t) in order to avoid detection, and then distort a delayed version of the stored signal, which may then have characteristics that are more easily detected, as, for instance, clipped noise. Thus the transmitted signal preserves its random nature. The correlation in the receiver may be done with simpler equipment, for instance, by clipping, and then eliminating the multiplier, as previously mentioned. Notice that we use separability only of the signal process in order to derive this simplified range-preserving correlator, and assume only that the additive noise is independent and otherwise arbitrary. In this application it is of no importance whether the sum process, s(t) + n(t), is separable or not.

We could, instead of distorting the stored delayed signal, distort the return waveform given by Eq. 35, and still preserve the shape of the signal correlation function and simplify our correlator. The average output of our correlator would be

$$\overline{s(t - \tau_2)f[As(t - \tau_1) + n(t)]} = \iiiint x_2 f(Ax_1 + y) p_s(x_1, x_2; \tau_1 - \tau_2) p_n(y) p(A)\, dx_1\, dx_2\, dy\, dA$$

$$= \rho_s(\tau_1 - \tau_2)\iiint x_1 f(Ax_1 + y) p_s(x_1) p_n(y) p(A)\, dx_1\, dy\, dA$$

Thus the average output is proportional to the signal correlation function, as desired. If a clipper were chosen as the nonlinear device f, we could eliminate the multiplier for a gating network in the correlator and obtain a simplified correlator for this arrangement also. However, this may be less advantageous than the previous one. If we deliberately clip the return waveform, and the signal-to-noise ratio is very small, the output of the clipper will be essentially the polarity of the noise voltage. We shall then ruin our signal-to-noise ratio at the output of the correlator, for, by clipping, we are almost completely destroying the part of the return waveform which is correlated with the delayed signal. Thus, although the correlator is simpler, its performance is poor. We should, therefore, clip the delayed signal instead and use it as a gating signal in a simplified correlator. The signal-to-ratio should be better.

Although the separable class of processes is useful for designing simplified correlators, it offers no advantages in the computation of the signal-to-noise ratio for the simplified correlators. The signal-to-noise ratio is a property of the particular statistics involved, and separability does not contain enough simplification for this computation.

3.8 CONNECTION OF THE PRESENT WORK WITH OTHER RESULTS

It is of interest to see how separability is connected with previous work on the invariance property. Barrett and Lampard (8) showed that for their class Λ of probability density functions, the invariance property held. Brown (9) then generalized the class Λ. We shall work with Brown's class of probability density functions and show that under Brown's hypotheses, the process is a separable one. For Brown's class, for the joint probability density function $p(x_1, t_1; x_2, t_2)$ (allowing nonstationary processes), we have

$$p(x_1, t_1; x_2, t_2) = p_1(x_1, t_1)p_2(x_2, t_2) \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{mn}(t_1, t_2)\theta_m^{(1)}(x_1, t_1)\theta_n^{(2)}(x_2, t_2) \tag{36}$$

where $p_1(x_1, t_1)$ and $p_2(x_2, t_2)$ are the first-order probability density functions of the processes $x_1(t)$ and $x_2(t)$, respectively. The θ's are polynomials satisfying

$$\int p_1(x, t)\theta_m^{(1)}(x, t)\theta_n^{(1)}(x, t)\, dx = \delta_{mn} \tag{37a}$$

$$\int p_2(x, t)\theta_m^{(2)}(x, t)\theta_n^{(2)}(x, t)\, dx = \delta_{mn} \tag{37b}$$

and a_{mn} is given by

$$a_{mn}(t_1, t_2) = \iint p(x_1, t_1; x_2, t_2)\theta_m^{(1)}(x_1, t_1)\theta_n^{(2)}(x_2, t_2)\, dx_1\, dx_2$$

Then

$$g(x_2; t_1, t_2) = \int \left[x_1 - \mu_1(t_1)\right] p(x_1, t_1; x_2, t_2)\, dx_1$$

$$= \sum_{n=0}^{\infty} a_{1n}(t_1, t_2)p_2(x_2, t_2)\theta_n^{(2)}(x_2, t_2)\sigma_1(t_1) \tag{38}$$

in which we have used Eqs. 36, 37, and the following equation of Barrett and Lampard (8),

$$\theta_1(x_1, t_1) = \frac{x_1 - \mu_1(t_1)}{\sigma_1(t_1)}$$

Brown has shown that if the joint probability density function can be expanded in a series such as Eq. 36, then

$$a_{1n}(t_1, t_2) = b_n(t_2)\rho(t_1, t_2) \qquad \text{for all } n \tag{39}$$

is a necessary and sufficient condition for the invariance property to hold. But substitution of this relation in formula 38 for $g(x_2; t_1, t_2)$ gives

$$\begin{aligned} g(x_2; t_1, t_2) &= \phi(t_1, t_2) \frac{p_2(x_2, t_2)}{\sigma_2(t_2)} \sum_{n=0}^{\infty} b_n(t_2)\theta_n^{(2)}(x_2, t_2) \\ &= \phi(t_1, t_2) g_1(x_2, t_2) \end{aligned} \tag{40}$$

which represents a separable process. Thus,

> Brown's class of processes (and therefore Barrett and Lampard's), for which $a_{1n}(t_1, t_2) = b_n(t_2)\rho(t_1, t_2)$ for all n, is a subset of the class of separable processes.

Note that under Brown's formulation, determination of Eq. 39 would be a formidable task indeed (some of the examples in Section VI will illustrate this). We would need to determine the orthogonal polynomials, the functions $a_{1n}(t_1, t_2)$, if they exist, and then see if the joint probability density function could be expanded in a double series (Eq. 36). Instead, all we need to do is to perform a single integral on the joint probability density function and test the separability of the resulting function.

We have shown that if Eq. 39 is true, the process is separable. Conversely, we can show that if the process is separable, and if the joint probability density function can be expanded in a double series (Eq. 36), then Eq. 39 is true if the left-hand side exists. We have

$$\begin{aligned} a_{1n}(t_1, t_2) &= \iint \theta_1^{(1)}(x_1, t_1)\theta_n^{(2)}(x_2, t_2)p(x_1, t_1; x_2, t_2)\, dx_1\, dx_2 \\ &= \frac{1}{\sigma_1(t_1)} \int g(x_2; t_1, t_2)\theta_n^{(2)}(x_2, t_2)\, dx_2 \\ &= \frac{\phi(t_1, t_2)}{\sigma_1(t_1)} \int g_1(x_2, t_2)\theta_n^{(2)}(x_2, t_2)\, dx_2 \\ &= \rho(t_1, t_2)\sigma_2(t_2) \int g_1(x_2, t_2)\theta_n^{(2)}(x_2, t_2)\, dx_2 \\ &= b_n(t_2)\rho(t_1, t_2) \end{aligned}$$

Now if we are dealing with a single process, our joint probability density function becomes the second-order probability density function of the process, and we can show some interesting facts about the sequence $\{b_n(t_2)\}$. Setting $t_1 = t_2 \equiv t$ in Eq. 40, and dropping superfluous subscripts and superscripts because we have only one process, we have

$$g(x_2; t, t) = \sigma(t)p(x_2, t) \sum_{n=0}^{\infty} b_n(t)\theta_n(x_2, t)$$

But

$$\begin{aligned} g(x_2; t, t) &= \int \left[x_1 - \mu(t)\right] p(x_1, t; x_2, t)\, dx_1 \\ &= \int \left[x_1 - \mu(t)\right] p(x_1, t)\delta(x_2 - x_1)\, dx_1 \\ &= \left[x_2 - \mu(t)\right] p(x_2, t) \end{aligned}$$

Therefore

$$\left[x_2 - \mu(t)\right] p(x_2, t) = p(x_2, t) \sum_{n=0}^{\infty} b_n(t)\sigma(t)\theta_n(x_2, t)$$

$$= p(x_2, t)\left[b_o(t)\sigma(t) + b_1(t)\{x_2 - \mu(t)\} + \sum_{n=2}^{\infty} b_n(t)\sigma(t)\theta_n(x_2, t)\right] \quad (41)$$

Now

$$\begin{aligned} a_{10}(t_1, t_2) &= \iint \theta_1(x_1, t_1)\theta_o(x_2, t_2)p(x_1, t_1; x_2, t_2)\, dx_1\, dx_2 \\ &= \int \theta_1(x_1, t_1)p(x_1, t_1)\, dx_1 = 0 \end{aligned}$$

since $\theta_o(x_2, t_2) = 1$, according to Barrett and Lampard (8). Also, since $a_{11}(t_1, t_2) = \rho(t_1, t_2)$, $b_1(t) = 1$. Therefore, Eq. 41 becomes

$$\left[x_2 - \mu(t)\right] p(x_2, t) = \left[x_2 - \mu(t)\right] p(x_2, t) + p(x_2, t) \sum_{n=0}^{\infty} b_n(t)\sigma(t)\theta_n(x_2, t)$$

Therefore

$$\sum_{n=2}^{\infty} b_n(t)\theta_n(x_2, t) = 0$$

and

$$b_n(t) = 0 \qquad \text{for } n \geqslant 2$$

This is Brown's class Λ^*. A familiar result,

$$g(x_2; t_1, t_2) = \phi(t_1, t_2)\, \frac{x_2 - \mu(t_2)}{\sigma(t_2)}\, p(x_2, t_2)$$

also follows. Thus, the requirement of Eq. 39 for one process results in the process being separable and all the b_n's being zero, except b_1 which equals one. No such simple relation in the sequence $\{b_n(t)\}$ exists for two processes.

Conversely, if a process is separable, then

$$a_{1n}(t_1, t_2) = \iint \theta_1(x_1, t_1)\theta_n(x_2, t_2)p(x_1, t_1; x_2, t_2)\, dx_1\, dx_2$$

$$= \frac{1}{\sigma(t_1)} \int \theta_n(x_2, t_2)g(x_2; t_1, t_2)\, dx_2$$

$$= \delta_{1n}\rho(t_1, t_2)$$

and

$$b_n(t) = \delta_{1n} \qquad \text{for all } n$$

Returning to the more general case of two processes, we have shown that separability is identical with the satisfaction of Eq. 39 whenever the quantities in Eq. 39 exist. Also, in Section I, we showed that separability is identical with cross moments, and hence

$$\overline{[x_1(t_1) - \mu_1(t_1)]x_2^n(t_2)} = b_n(t_2)\phi(t_1, t_2) \qquad \text{for all n.}$$

Therefore there must be a direct connection between the cross moments and the quantities $a_{1n}(t_1, t_2)$. We shall now derive the connection.

$$a_{1n}(t_1, t_2) = \iint \theta_1^{(1)}(x_1, t_1)\theta_n^{(2)}(x_2, t_2)p(x_1, t_1; x_2, t_2)\, dx_1\, dx_2$$

and

$$\theta_n^{(2)}(x_2, t_2) = \sum_{k=0}^{n} a_k^{(n)}(t_2)x_2^k$$

Therefore

$$a_{1n}(t_1, t_2) = \iint p(x_1, t_1; x_2, t_2)\frac{x_1 - \mu_1(t_1)}{\sigma_1(t_1)} \sum_{k=0}^{n} a_k^{(n)}(t_2)x_2^k\, dx_1\, dx_2$$

$$= \sum_{k=0}^{n} a_k^{(n)}(t_2)\, \frac{\overline{x_1(t_1)x_2^k(t_2)} - \overline{x_1(t_1)}\,\overline{x_2^k(t_2)}}{\sigma_1(t_1)} \qquad (42)$$

Now if

$$\overline{x_1(t_1)x_2^k(t_2)} - \overline{x_1(t_1)}\,\overline{x_2^k(t_2)} = d_k(t_2)\phi(t_1, t_2) \qquad \text{for all k}$$

then

$$a_{1n}(t_1, t_2) = b_n(t_2)\rho(t_1, t_2) \qquad \text{for all n}$$

Conversely, if

$$a_{1n}(t_1, t_2) = b_n(t_2)\rho(t_1, t_2) \qquad \text{for all } n$$

then, taking successively n = 0, 1, 2, ..., in Eq. 42, we find that

$$\overline{x_1(t_1)x_2^k(t_2)} - \overline{x_1(t_1)}\,\overline{x_2^k(t_2)} = d_k(t_2)\phi(t_1, t_2) \qquad \text{for all } k \tag{43}$$

Thus an alternative statement of Brown's condition for the invariance property to hold can be made in terms of cross moments, as in Eq. 43. Now the determination of the satisfaction of Eq. 43 is much simpler than that for Eq. 39, because Eq. 39 requires the determination of the orthogonal polynomials associated with the process and the functions $a_{1n}(t_1, t_2)$. Also, expression 43 has a ready physical interpretation, whereas Eq. 39 does not. On the other hand, both of these approaches suffer in the cases in which higher-order moments do not exist, but there are no such troubles with separability.

3.9 DOUBLE NONLINEAR TRANSFORMATIONS

A further variation that bears investigation is that of inserting in the top lead of Fig. 1 another nonlinear no-memory network f'. What, then, is the necessary and sufficient condition that must be imposed on the input statistics for the input and output crosscovariance functions to be proportional? We shall call it the invariance property for this more general network. By an approach that is strictly analogous to that in section 3.4, we find, for the stationary case, that

$$p(x_1, x_2; \tau) - p_1(x_1)p_2(x_2) = h(x_1; x_2)\phi(\tau) \tag{44}$$

is a necessary and sufficient condition for the invariance property to hold. The function $h(x_1, x_2)$ is a function only of the variables x_1 and x_2. Examples of processes that satisfy Eq. 44 are difficult to find, although this class is not empty. A trivial example of processes in this class is $x_1(t) = C_1$ and $x_2(t) = C_2$. Luce (7) proved the sufficiency of Eq. 44 but was unable to prove its necessity, except under very restrictive conditions. Extensions to nonstationary cases are straightforward.

IV. EXTENSION OF BOOTON'S EQUIVALENT GAIN NETWORKS

4.1 SEPARABILITY AND EQUIVALENT GAIN TECHNIQUES

In some problems of electrical engineering the statistical analysis of networks is exceedingly difficult. A simpler approximate method of analyzing networks was proposed and investigated by Booton (11, 12, 13). In Booton's method, a nonlinear no-memory device was replaced by a linear no-memory device, so chosen that the mean-square difference between the two outputs, for a Gaussian input, was minimum. By this method, the analysis of feedback networks, for instance, with single nonlinearities, is possible, since the techniques of linear feedback networks are well established and are relatively easy to apply.

In this section we extend Booton's results to the (stationary) separable class of inputs and allow more general linear networks. Let x(t) be the input process into a nonlinear no-memory network characterized by a function f. Then the output y(t) is

$$y(t) = f[x(t)]$$

We now want to replace the nonlinear no-memory network f by a realizable linear memory-capable network plus a constant, with output

$$r(t) = C + \int_0^\infty h(\alpha)x(t-\alpha)\, d\alpha$$

in such a way that the mean-square error

$$E = \overline{[y(t) - r(t)]^2}$$

is minimum. By allowing a choice for the quantity C, and for the impulse response h(t), we can reduce the error E to a smaller value than with just an impulse response alone. Furthermore, the computations are no more difficult. The error is

$$E = \overline{\left[f[x(t)] - C - \int_0^\infty h(\alpha)x(t-\alpha)\, d\alpha\right]^2}$$

Then $\partial E/\partial C = 0$ gives

$$\overline{f[x(t)]} = C + \int_0^\infty h(\alpha)\overline{x(t-\alpha)}\, d\alpha$$

Substituting this value for C in the expression for the error E, we obtain

$$E = \overline{\left[f[x(t)] - \overline{f[x(t)]} - \int_0^\infty h(\alpha)[x(t-\alpha) - \overline{x(t-\alpha)}]\, d\alpha\right]^2}$$

We now have to choose $h(\alpha)$ in such a way as to minimize E. If we define a desired output d(t) as

$$d(t) = f[x(t)] - \overline{f[x(t)]}$$

and an input i(t) as

$$i(t) = x(t) - \overline{x(t)}$$

we have to minimize

$$E = \overline{\left[d(t) - \int_0^\infty h(\alpha)i(t-\alpha)\, d\alpha\right]^2}$$

by choice of $h(\alpha)$. According to Wiener (1), the optimum impulse response $h_o(\alpha)$ satisfies the integral equation

$$\phi_{id}(\tau) = \int_0^\infty h_o(\alpha)\phi_{ii}(\tau - \alpha)\, d\alpha, \quad \tau \geq 0 \tag{45}$$

Now

$$\phi_{ii}(\tau) = \overline{\left[x(t) - \overline{x(t)}\right]\left[x(t+\tau) - \overline{x(t+\tau)}\right]} = \sigma^2 \rho(\tau)$$

and

$$\phi_{id}(\tau) = \overline{\left[x(t) - \overline{x(t)}\right]\left[f[x(t+\tau)] - \overline{f[x(t+\tau)]}\right]}$$

$$= \overline{\left[x(t) - \overline{x(t)}\right] f\left[x(t+\tau)\right]}$$

$$= \iint (x_1 - \mu) f(x_2) p(x_1, x_2; \tau)\, dx_1\, dx_2 \tag{46}$$

Now let us suppose that the input process is a separable process. Then

$$\phi_{id}(\tau) = \int f(x_2)(x_2 - \mu)p(x_2)\, dx_2\, \rho(\tau)$$

Substitution of these expressions in the integral equation (Eq. 45) for $h_o(\alpha)$ gives

$$\rho(\tau) \int f(x_2)(x_2 - \mu)p(x_2)\, dx_2 = \int_0^\infty h_o(\alpha)\sigma^2 \rho(\tau - \alpha)\, d\alpha$$

for $\tau \geq 0$. Therefore

$$h_o(\alpha) = \int f(x) \frac{x - \mu}{\sigma^2} p(x)\, dx\, \delta(\alpha)$$

That is, the optimum mean-square linear memory-capable network which replaces a nonlinear no-memory network, has, in fact, no memory for a separable input process! It is merely an attenuator.

Thus, when Booton restricted himself to networks with no memory, he was choosing the best network out of the class of linear networks, for separable input processes.

In this case, the output of the approximating network becomes

$$r(t) = \overline{f[x(t)]} + \frac{1}{\sigma^2}\, \overline{f[x(t)]\left[x(t) - \overline{x(t)}\right]}\left[x(t) - \overline{x(t)}\right]$$

$$= \int f(x)p(x)\, dx + \frac{1}{\sigma^2} \int f(x)(x-\mu)p(x)\, dx\, \left[x(t)-\overline{x(t)}\right]$$

Thus, the dc and ac components of the output are separated. The dc c
output should give little trouble in the analysis of a linear system, and t
is a mere attenuator. Notice also that the constant relating the ac compc
output to the ac component of the input to the approximating network is

$$\frac{1}{\sigma^2} \int f(x)(x-\mu)p(x)\, dx = C_f$$

as given by Eq. 21. Thus the equivalent ac gain is the same as the constant of p
tionality in the invariance property. We have, then

$$r(t) = \overline{f[x(t)]} + C_f[x(t)-\overline{x(t)}]$$

Now let us define an error process for the approximation procedure as

$$e(t) = y(t) - r(t)$$

$$= f[x(t)] - \overline{f[x(t)]} - C_f[x(t)-\overline{x(t)}] \tag{47}$$

Then the input-error crosscorrelation function (or crosscovariance function, since $\overline{e(t)} = 0$) is

$$\phi_{xe}(\tau) = \overline{x(t)e(t+\tau)} = \overline{[x(t)-\mu]e(t+\tau)}$$

$$= \overline{[x(t)-\mu]\, \{f[x(t+\tau) - \overline{f[x(t+\tau)]} - C_f[x(t+\tau) - \overline{x(t+\tau)}]\}}$$

$$= \overline{[x(t)-\mu]f[x(t+\tau)]} - C_f\overline{[x(t)-\mu]x(t+\tau)}$$

$$= C_f\phi(\tau) - C_f\phi(\tau) = 0 \qquad \text{for all } \tau$$

Then the input and error processes are linearly independent for separable processes. The linear approximating network has done its best in attempting to resemble the nonlinear device and leaves only an error that is linearly independent of the input, if the input is a separable process.

We can also show that, for arbitrary nonlinear no-memory devices, the input and error are uncorrelated <u>only</u> if the input process is separable. To prove this, we suppose that

$$\phi_{xe}(\tau) = \overline{[x(t)-\mu]e(t+\tau)} = 0 \qquad \text{for all } \tau$$

for any f; the error is found by replacing the nonlinear device f by the optimum ear network. Then e(t) is given by Eq. 47. Therefore

$$0 = \overline{\left[x(t)-\overline{x(t)}\right]\left[f[x(t+\tau)] - \overline{f[x(t+\tau)]} - C_f\left[x(t+\tau) - \overline{x(t+\tau)}\right]\right]}$$

for any f, τ, with C_f given by Eq. 21. Therefore

$$0 = \overline{\left[x(t)-\overline{x(t)}\right]f[x(t+\tau)]} - C_f\overline{\left[x(t)-\overline{x(t)}\right]x(t+\tau)}$$

$$= \int [f(x_2) - C_f x_2]g(x_2, \tau)\, dx_2$$

$$= \int f(x_2)g(x_2, \tau)\, dx_2 - C_f\sigma^2 \rho(\tau)$$

$$= \int f(x_2)g(x_2, \tau)\, dx_2 - \int f(x_2)(x_2 - \mu)p(x_2)\, dx_2 \rho(\tau)$$

$$= \int f(x_2)[g(x_2, \tau) - (x_2 - \mu)p(x_2)\rho(\tau)]\, dx_2$$

for any f, τ. Therefore

$$g(x_2, \tau) = (x_2 - \mu)p(x_2)\rho(\tau)$$

and we have a separable input process. Thus we have shown that

separability is a necessary and sufficient condition for the input and error in the approximation scheme to be uncorrelated for arbitrary nonlinear no-memory devices.

Let us now look at the output autocovariance functions of the approximating linear network and the actual nonlinear device. We have

$$\phi_{yy}(\tau) = \overline{\left[f[x(t)] - \overline{f[x(t)]}\right]\left[f[x(t+\tau)] - \overline{f[x(t+\tau)]}\right]}$$

$$= \overline{\left[C_f\left[x(t)-\overline{x(t)}\right] + e(t)\right]\left[C_f\left[x(t+\tau) - \overline{x(t+\tau)}\right] + e(t+\tau)\right]}$$

$$= C_f^2\phi(\tau) + C_f[\phi_{xe}(\tau) + \phi_{xe}(-\tau)] + \phi_{ee}(\tau)$$

But if our input process is separable, we have

$$\phi_{xe}(\tau) = \phi_{xe}(-\tau) = 0 .$$

for all τ. Then

$$\phi_{yy}(\tau) = C_f^2\phi(\tau) + \phi_{ee}(\tau)$$

Now

$$\phi_{rr}(\tau) = \overline{\left[r(t)-\overline{r(t)}\right]\left[r(t+\tau) - \overline{r(t+\tau)}\right]}$$

$$= C_f^2\overline{[x(t)-\overline{x(t)}]\,[x(t+\tau) - \overline{x(t+\tau)}]}$$

$$= C_f^2\phi(\tau)$$

Therefore

$$\phi_{yy}(\tau) = \phi_{rr}(\tau) + \phi_{ee}(\tau)$$

This relation holds true only for separable processes. In general, we have to include cross-product terms. Thus, neglecting the error term in the linear approximation to f amounts to neglecting $\phi_{ee}(\tau)$ in the output autocovariance function. The formula given above may be useful in measuring how good the approximation is. For, we have have

$$G_y(f) = G_r(f) + G_e(f)$$

where $G_z(f)$ is the frequency spectrum of the z(t) process. Then if we compute $G_r(f)$ and $G_e(f)$, and inspect the transfer functions following the nonlinear device, we can tell how much of $G_e(f)$ would be suppressed in comparison with $G_r(f)$, and obtain a rough idea of how well our over-all approximating system will perform.

We have minimized $E = \phi_{ee}(0)$ to obtain the optimum linear network. For separable processes, we in fact obtain identically the same linear network by minimizing $\phi_{ee}(\tau)$ for any fixed τ.

Also, since

$$r(t) - \overline{r(t)} = C_f\left[x(t)-\overline{x(t)}\right]$$

the approximating response-error crosscorrelation function (or crosscovariance function, since $\overline{e(t)} = 0$) is

$$\phi_{re}(\tau) = \overline{r(t)e(t+\tau)} = \overline{\left[r(t)-\overline{r(t)}\right]e(t+\tau)}$$

$$= C_f\overline{[x(t)-\overline{x(t)}]e(t+\tau)}$$

$$= C_f\phi_{xe}(\tau) = 0 \qquad \text{for all } \tau$$

for a separable input process. Thus the response is also linearly independent of the error for a separable input process.

For other processes than separable processes, the optimum linear memory-capable network impulse response will not be a delta function. In general, from Eq. 45, we have

$$\phi_{id}(\tau) = \int_0^{\infty} h_o(\alpha)\phi_{ii}(\tau-\alpha)\, d\alpha, \qquad \tau \geq 0$$

Now

$$\phi_{id}(\tau) = \int f(x_2)g(x_2, \tau)\, dx_2$$

from Eq. 46, and

$$\phi_{ii}(\tau) = \int x_2 g(x_2, \tau)\, dx_2 = \sigma^2 \rho(\tau)$$

Therefore

$$\int f(x_2)g(x_2, \tau)\, dx_2 = \sigma^2 \int_0^{\infty} h_o(\alpha)\rho(\tau-\alpha)\, d\alpha \tag{48}$$

for $\tau \geq 0$. Now in order for $h_o(\alpha)$ to be a delta function, the integral on the left-hand side of Eq. 48 must become a constant multiplied by $\rho(\tau)$. Since this happens only coincidentally (for some particular functions f) for other processes than separable processes, we see that the optimum linear network, in general, contains memory. Thus for the separable class alone do we get such an easy approximation network. Let it be noted that although the solution of Eq. 48 is not difficult, in order to use this linear network in a feedback network, for instance, we would probably have to find the system transfer function. But we are not assured that the system transfer function will end up as a rational function, and it is for rational transfer functions that our computations and manipulations become very simple. If we wanted a rational transfer function, we would have a further approximation problem. But for the separable class of processes, we are always assured that the transfer function will be rational; in fact, it will be a constant.

The error in a particular approximation can be evaluated fairly easily for the separable class:

$$E = \overline{\left[f[x(t)] - \overline{f[x(t)]} - C_f[x(t)-\overline{x(t)}]\right]^2}$$

$$= \overline{\left\{f[x(t)] - \overline{f[x(t)]}\right\}^2} - 2C_f\overline{\left[x(t)-\overline{x(t)}\right]f[x(t)]} + C_f^2\sigma^2$$

$$= \overline{f^2[x(t)]} - \overline{f[x(t)]}^2 - C_f^2\sigma^2$$

or

$$E = \int f^2(x)p(x)\, dx - \left[\int f(x)p(x)\, dx\right]^2 - \left[\int f(x)\, \frac{x-\mu}{\sigma}\, p(x)\, dx\right]^2$$

and only the first-order probability density function is needed.

4.2 NONLINEAR NETWORKS OF GREATER GENERALITY

A variation of this problem is possible, in which the output of the nonlinear network is a combination of a nonlinear no-memory operation and a linear unrealizable operation on the input:

$$y(t) = f[x(t)] + \int u(\alpha)x(t-\alpha)\, d\alpha$$

The problem is to approximate to y(t) by a linear realizable network. By an approach analogous to that given in the beginning of Section IV, we find that the optimum impulse response for a separable input process is given by

$$h_o(\alpha) = \int f(x)\frac{x-\mu}{\sigma^2}\, p(x)\, dx\ \delta(\alpha) + h^*(\alpha)$$

where $h^*(\alpha)$ satisfies the integral equation

$$\int_0^\infty h^*(\alpha)\rho(\tau-\alpha)\, d\alpha = \int u(\alpha)\rho(\tau-\alpha)\, d\alpha$$

for $\tau \geqslant 0$. Thus, for a separable process, the part of the optimum impulse response corresponding to the nonlinear no-memory operation has no memory. The remaining part is approximated to in the usual fashion (1).

Another class of nonlinear networks that is easy to deal with, if the input process is separable, is the class (14, 15) whose output is given by

$$y(t) = \int_0^\infty F[\alpha, x(t-\alpha)]\, d\alpha \tag{49}$$

where x(t) is the input. This class includes realizable linear networks and nonlinear no-memory networks as special cases. We wish to approximate to y(t) by a realizable linear network (let $\overline{x(t)} = 0$ for convenience), as follows:

$$r(t) = \int_0^\infty h(\alpha)x(t-\alpha)\, d\alpha$$

Then, according to Wiener (1), we have to solve

$$\phi_{id}(\tau) = \int_0^\infty h_o(\alpha)\phi_{ii}(\tau-\alpha)\, d\alpha$$

for $\tau \geqslant 0$, where

$$\phi_{ii}(\tau) = \overline{x(t)x(t+\tau)} = \sigma^2\rho(\tau)$$

and

$$\phi_{id}(\tau) = \overline{x(t)\int_0^\infty F[\alpha, x(t+\tau-\alpha)]\, d\alpha} =$$

$$= \int_0^\infty \iint x_1 F(\alpha, x_2) p(x_1, x_2; \tau-\alpha)\, dx_1\, dx_2\, d\alpha$$

$$= \int_0^\infty \int x_2 F(\alpha, x_2) p(x_2) \rho(\tau-\alpha)\, dx_2\, d\alpha$$

in which we have used the fact that the input process is separable (with zero mean). Therefore

$$\int_0^\infty \rho(\tau-\alpha)\left[\int x_2 F(\alpha, x_2) p(x_2)\, dx_2\right] d\alpha = \sigma^2 \int_0^\infty h_o(\alpha)\rho(\tau-\alpha)\, d\alpha$$

for $\tau \geqslant 0$. Therefore

$$h_o(\alpha) = \int \frac{x}{\sigma^2} F(\alpha, x) p(x)\, dx$$

for $\alpha \geqslant 0$. Thus,

when the input process is separable (with zero mean) and the actual nonlinear operation is given by

$$\int_0^\infty F[\alpha, x(t-\alpha)]\, d\alpha$$

the optimum linear impulse response is easily found as

$$h_o(\alpha) = \int \frac{x}{\sigma^2} F(x, \alpha) p(x)\, dx$$

No integral equation for $h_o(\alpha)$ need be solved.

However this optimum linear network, in general, contains memory.

Extensions to time-variant statistics are straightforward by using the results of previous sections.

In an actual <u>feedback</u> network, it may be very difficult, if not impossible, to tell whether or not the input into a nonlinear device is a separable process. This requires additional work, and can perhaps be best investigated by use of the moment formula,

$$\overline{x(t)x^n(t+\tau)} - \overline{x(t)}\,\overline{x^n(t+\tau)} = b_n \rho(\tau) \qquad \text{for all n}$$

for separability to hold. The determination of the second-order probability density function or characteristic function is out of the question.

V. MARKOV PROCESSES AND BEST ESTIMATE PROCEDURE

5.1 CORRELATION FUNCTIONS OF MARKOV PROCESSES

Doob (16) has shown that a Gaussian (zero-mean) Markov process of order one has a correlation function that must be an exponential function. Barrett and Lampard (8) showed that any Markov process of order one in their class Λ must also have an exponential correlation function. We now demonstrate that this holds also for the (stationary) separable class of random processes which is a Markov process of first order. We assume a zero-mean process for convenience. We have, since the process is Markov of first order,

$$p(x_1;x_2,\tau_1;x_3,\tau_2) = \frac{p(x_1,x_2;\tau_1)p(x_2,x_3;\tau_2)}{p(x_2)}$$

where $\tau_1, \tau_2 \geqslant 0$; τ_1 is the time between samples x_1 and x_2; and τ_2 is the time between samples x_2 and x_3. Then

$$p(x_1,x_3;\tau_1+\tau_2) = \int \frac{p(x_1,x_2;\tau_1)p(x_2,x_3;\tau_2)}{p(x_2)}\, dx_2$$

and with the use of separability, we have

$$\begin{aligned}\int x_1 p(x_1,x_3;\tau_1+\tau_2)\, dx_1 &= x_3 p(x_3)\rho(\tau_1+\tau_2)\\ &= \int x_1 \int p(x_1)p(x_2|x_1;\tau_1)p(x_3|x_2;\tau_2)\, dx_2\, dx_1\\ &= \int p(x_3|x_2;\tau_2)\int x_1 p(x_1,x_2;\tau_1)\, dx_1\, dx_2\\ &= \int p(x_3|x_2;\tau_2)x_2 p(x_2)\rho(\tau_1)\, dx_2\\ &= \rho(\tau_1)\int x_2 p(x_2,x_3;\tau_2)\, dx_2\\ &= \rho(\tau_1)x_3 p(x_3)\rho(\tau_2)\end{aligned}$$

Therefore

$$\rho(\tau_1+\tau_2) = \rho(\tau_1)\rho(\tau_2), \qquad \tau_1, \tau_2 \geqslant 0$$

The only continuous real solution to this equation is

$$\rho(\tau) = e^{a\tau}$$

but, since $|\rho(\tau)| \leqslant \rho(0) = 1$, we have

$$\rho(\tau) = e^{-C|\tau|} \qquad \text{for all } \tau$$

where $C \geqslant 0$. Thus

a separable (zero-mean) Markov process of order one has a correlation function that is an exponential.

We can also show that a separable Markov process of order one has a second-order correlation function that is an exponential. The second-order correlation function is

$$\overline{x(t)x(t+\tau_1)x(t+\tau_1+\tau_2)} = \iiint x_1x_2x_3p(x_1;x_2,\tau_1;x_3,\tau_2)\,dx_1\,dx_2\,dx_3$$

$$= \int x_2 \frac{1}{p(x_2)} \int x_1p(x_1,x_2;\tau_1)\,dx_1 \int x_3p(x_2,x_3;\tau_2)\,dx_3\,dx_2$$

$$= \int x_2 \frac{1}{p(x_2)}\, x_2p(x_2)\rho(\tau_1)x_2p(x_2)\rho(\tau_2)\,dx_2$$

$$= \rho(\tau_1+\tau_2) \int x_2^3p(x_2)\,dx_2$$

$$= \overline{x^3(t)}\, e^{-C|\tau_1+\tau_2|}$$

Thus

the second-order correlation function is an exponential for separable Markov processes of order one. Also, the location of the middle sample in the definition of the correlation function is unimportant.

We see this readily by setting $\alpha_1 = \tau_1$, $\alpha_2 = \tau_1 + \tau_2$. Then

$$\overline{x(t)x(t+\alpha_1)x(t+\alpha_2)} = \overline{x^3(t)}\, e^{-C|\alpha_2|}$$

and α_1 does not appear. However, this formula holds only for $0 \leqslant \alpha_1 \leqslant \alpha_2$.

Attempts to use separability to find the form of the third-order correlation function, for Markov processes of order one, met with failure because of lack of knowledge of statistics. We need knowledge about

$$\int x_1^2p(x_1,x_2;\tau)\,dx_1$$

about which we have made no assumptions.

5.2 SINGLE SAMPLE BEST ESTIMATE PROCEDURE

Given the past history of a signal plus noise, it is sometimes of interest to find out what the value of the signal will be at some future time. This is a combined filtering and prediction problem. When the signal and noise are random processes, we cannot expect to predict exactly what the signal will be at some future time; we can only hope to get an estimate of the signal. We define an error and attempt to minimize it by choice of an operator on the past history of the signal plus noise. When the error is defined as the mean-square difference between the actual future value of the signal and the imperfectly predicted value, we find that the best operator on the history of signal plus noise is the conditional mean of the signal at the future time, given the past history of the signal plus noise. This is a well-known result, but is often an ideal approach that cannot be realized practically because of limited knowledge of the statistics of the input.

Instead of using the complete past history of the signal plus noise, we often restrict ourselves to using a few past samples of the signal plus noise. This simplifies the statistics that we need to know in order to find the best estimate network. In this report we restrict ourselves to using only one sample of the (stationary) signal plus noise, the present value, and attempt to predict the signal at some time τ seconds in the future. Let

$$y(t) = s(t) + n(t)$$

Then the best estimate or operator on y $[= y(t)]$ is

$$e(t+\tau) = \int s p(s \mid y;\tau)\, ds$$

where $p(s \mid y;\tau)$ is the conditional probability density function of the signal at time $t + \tau$, given the value of the signal plus noise at time t. Now

$$p(s \mid y;\tau) = \frac{p(s, y;\tau)}{p(y)}$$

and

$$p(s, y;\tau) = \int p_s(s_1, s;\tau) p_n(y - s_1)\, ds$$

where $p_s(s_1, s;\tau)$ is the second-order probability density function of the signal, $p_n(n)$ is the first-order probability density function of the noise, and $p(y)$ is the first-order probability density function of the sum of signal and noise. Then

$$e(t+\tau) = \frac{1}{p(y)} \iint s p_s(s_1, s;\tau) p_n(y - s_1)\, ds\, ds_1$$

Now let us assume that the signal process is a separable process. Then

$$\int sp_s(s_1, s;\tau)\, ds = p_s(s_1)[(s_1 - \mu_s)\rho_s(\tau) + \mu_s]$$

and

$$e(t+\tau) = \frac{1}{p(y)} \int p_n(y - s_1)p_s(s_1)[(s_1 - \mu_s)\rho_s(\tau) + \mu_s]\, ds_1$$

Now

$$p_s(s_1)p_n(y - s_1) = p(s_1, s_1 + n = y)$$

Therefore

$$\frac{1}{p(y)}\, p_n(y - s_1)p_s(s_1) = p(s_1 | s_1 + n = y)$$

and

$$e(t+\tau) = \int p(s_1 | s_1 + n = y)[(s_1 - \mu_s)\rho_s(\tau) + \mu_s]\, ds_1$$

$$= \int (s_1 - \mu_s)p(s_1 | s_1 + n = y)\, ds_1\, \rho_s(\tau) + \mu_s$$

also

$$e(t) = \int (s_1 - \mu_s)p(s_1 | s_1 + n = y)\, ds_1 + \mu_s \tag{50}$$

Therefore

$$e(t+\tau) = e(t)\rho_s(\tau) + \mu_s[1 - \rho_s(\tau)]$$

Now $e(t)$ is the best estimate of the signal, given the value of the signal plus noise at the same instant of time. Thus it is the optimum filter. Hence

> for prediction and filtering of a separable signal in arbitrary additive independent noise, the best operator on one sample is an optimum filter with a gain $\rho_s(\tau)$ and an additive constant $\mu_s[1 - \rho_s(\tau)]$.

The optimum filter is found by substituting the first-order statistics of the signal and noise in Eq. 50. In addition to these first-order statistics, we need know only the normalized covariance function $\rho_s(\tau)$ of the signal process, if it is a separable process. Such a simple relation does not hold true for other than separable processes.

In the special case in which there is no noise, the optimum prediction is

$$e(t+\tau) = s(t)\rho_s(\tau) + \mu_s[1 - \rho_s(\tau)]$$

which is a gain factor multiplied by the present value with an additive constant. This linear relation, known to be true for Gaussian processes, is thus seen to be true for separable processes. The error with the latter process is easily evaluated:

$$E = \overline{[e(t+\tau) - s(t+\tau)]^2}$$

$$= \overline{[s(t)\rho_s(\tau) + \mu_s[1 - \rho_s(\tau)] - s(t+\tau)]^2}$$

$$= \sigma^2\left[1 - \rho_s^2(\tau)\right] + 2\mu_s^2\rho_s(\tau)\,[1 - \rho_s(\tau)]$$

From the results under separable Markov processes of order one with zero mean, we see that the optimum predictor with the use of one sample of (Markov) signal gives as its output

$$e(t+\tau) = s(t)\,e^{-C|\tau|}$$

Thus, a mere exponential attenuator is the best predictor.

The present results are easily extended to nonstationary statistics.

VI. EXAMPLES OF SEPARABLE PROCESSES

6.1 GAUSSIAN PROCESS

$$p(x_1, x_2; \tau) = \left[2\pi\sigma^2(1 - \rho^2(\tau))^{1/2}\right]^{-1} \exp\left[- \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 - 2\rho(\tau)(x_1 - \mu)(x_2 - \mu)}{2\sigma^2[1 - \rho^2(\tau)]}\right]$$

$$g(x, \tau) = (x-\mu) \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rho(\tau)$$

$$G(\xi, \tau) = -\sigma^2 \xi\ e^{-\frac{\sigma^2}{2}\xi^2} \rho(\tau)$$

6.2 SINE-WAVE PROCESS

$$x(t) = A \cos(\omega t + \theta), \ p(\theta) = \frac{1}{2\pi}, \quad 0 < \theta < 2\pi$$

$$f(\xi_1, \xi_2; \tau) = J_o\left[A(\xi_1^2 + \xi_2^2 + 2\xi_1\xi_2 \cos \omega\tau)^{1/2}\right]$$

$$G(\xi, \tau) = - AJ_1(A\xi) \cos \omega\tau$$

$$g(x, \tau) = \begin{cases} 0, & |x| > A \\ \dfrac{x}{\pi\sqrt{A^2 - x^2}} \cos \omega\tau, & |x| < A \end{cases}$$

6.3 ENVELOPE OF SQUARED NARROW-BAND GAUSSIAN PROCESS

$$p(x_1, x_2; \tau) = \frac{\gamma}{4\sigma^2} I_o[\mu\gamma\sqrt{x_1 x_2}]\, e^{-\frac{\gamma}{2}(x_1+x_2)}, \quad x_1, x_2 \geq 0$$

where σ^2 is the power of the Gaussian process (8), $w(f)$ is its spectrum (defined for $f \geq 0$),

$$\mu^2 = \mu^2(\tau) = \frac{\int_0^\infty \int_0^\infty w(f_1)w(f_2) \cos 2\pi(f_1 - f_2)\tau \, df_1 \, df_2}{\left[\int_0^\infty w(f)\, df\right]^2}$$

and

$$\gamma = \gamma(\tau) = [\sigma^2(1 - \mu^2(\tau))]^{-1}$$

Then

$$g(x, \tau) = \frac{x - 2\sigma^2}{2\sigma^2} \mu^2(\tau) e^{-\frac{x}{2\sigma^2}}, \qquad x \geq 0$$

$$G(\xi, \tau) = f'(\xi)\mu^2(\tau)$$

where

$$f(\xi) = \frac{e^{-j2\sigma^2\xi}}{1 - j2\sigma^2\xi}$$

6.4 SQUARED GAUSSIAN PROCESS

$$G(\xi, \tau) = -\frac{2\sigma^4\xi e^{-j\xi\sigma^2}}{(1 - j2\sigma^2\xi)^{3/2}} \rho^2(\tau)$$

where $\rho(\tau)$ is the normalized covariance function of the zero-mean Gaussian process, and σ^2 is its power. It is easily checked that no other power of a Gaussian process yields a separable process.

6.5 SQUARE-WAVE ALTERNATING BETWEEN a AND b, RANDOMLY OR OTHERWISE

Let $\overline{x(t)} = \mu$. Now

$$x^n(t) = a^n + \frac{x(t) - a}{b - a}(b^n - a^n)$$

Therefore

$$\overline{[x(t)-\mu]x^n(t+\tau)} = \overline{[x(t)-\mu]x(t+\tau)}\,\frac{b^n - a^n}{b - a}$$

and the process is separable with a scale factor in the cross moments of value

$$\frac{b^n - a^n}{b - a} = \sum_{k=1}^{n} b^{k-1}a^{n-k} \qquad (n \geq 1)$$

6.6 CARRIER-SUPPRESSED AMPLITUDE-MODULATED PROCESS

$$y(t) = x(t) \sin(\omega t + \theta)$$

If the $x(t)$ process is separable and independant of the uniformly distributed random variable θ, the $y(t)$ process is recognized as being a product of two independent processes. Since the sine wave is a separable process with zero mean, $y(t)$ will be a separable process if $x(t)$ is a separable process with a zero mean. (Actually, if $x(t)$ has a nonzero mean, it appears as a frequency component at the carrier and is eliminated.)

If we do not suppress the carrier, we have

$$y(t) = [1 + x(t)] \sin (\omega t + \theta)$$

and

$$G(\xi, \tau) = \overline{jy(t)\, e^{j\xi y(t+\tau)}}$$

$$= -\beta_1(\xi) \cos \omega\tau - \beta_2(\xi) \cos \omega\tau\rho(\tau)$$

where

$$\beta_1(\xi) = \int J_1[\xi(1+x)]p(x)\, dx$$

and

$$\beta_2(\xi) = \int J_1[\xi(1+x)]xp(x)\, dx$$

Here $p(x)$ and $\rho(\tau)$ are the first-order probability density function and the normalized covariance function of the $x(t)$ process, respectively. Thus $y(t)$ is not a separable process. However, we can generalize the definition of separability (as we do in Section VII) to include cases similar to this one.

6.7 PHASE- (OR FREQUENCY-) MODULATED PROCESS

$$y(t) = A \cos [\omega\tau + \theta + z(t)]$$

where $z(t)$ is independent of the uniformly distributed random variable θ. We let $\overline{z(t)} = 0$ with no loss of generality, since an additive constant on θ changes nothing. This form represents a frequency modulated wave, as well as a phase modulated wave, since $z(t)$ itself may be an integral of another random process. Then, since $\overline{y(t)} = 0$,

$$G(\xi, \tau) = \overline{jy(t)\, e^{j\xi y(t+\tau)}}$$

$$= -A\, \overline{\cos [\omega\tau - \{z(t+\tau) - z(t)\}]}\, J_1(A\xi)$$

which is a separable function. Notice that we have assumed nothing about the separability of the $z(t)$ process. Thus

> a phase- (or frequency-) modulated process is separable <u>regardless of the modulation</u>.

Now

$$\overline{\cos\left[\omega\tau - \{z(t+\tau) - z(t)\}\right]}$$

$$= \overline{\mathrm{Re}\left\{e^{j\left[\omega\tau - \{z(t+\tau) - z(t)\}\right]}\right\}}$$

$$= \mathrm{Re}\left\{e^{j\omega\tau}\overline{e^{j[z(t) - z(t+\tau)]}}\right\}$$

$$= \mathrm{Re}\ [e^{j\omega\tau}f_z(1, -1; \tau)]$$

and we need to know the second-order characteristic function of the z(t) process at two points. For the special case of a Gaussian modulation,

$$f_z(1, -1; \tau) = \exp[-\sigma^2(1-\rho(\tau))]$$

and therefore, in this case,

$$G(\xi, \tau) = -AJ_1(A\xi) \cos \omega\tau \exp[-\sigma^2(1-\rho(\tau))]$$

6.8 SQUARE-WAVE AND ARBITRARY PROCESS

We shall give a simple example of a process that is separable with respect to another process. Bussgang (6) has given one also. Let $x_1(t)$ be an arbitrary random process with zero mean, and let $x_2(t)$ be a square wave alternating between A and -A, randomly (or otherwise). Then

$$G(\xi, \tau) = \overline{j\left[x_1(t)-\overline{x_1(t)}\right] e^{j\xi x_2(t+\tau)}}$$

$$= -\frac{\sin \xi A}{A} \phi(\tau)$$

where $\phi(\tau)$ is the crosscorrelation function between $x_1(t)$ and $x_2(t)$.

6.9 REMARKS ON OTHER EXAMPLES

By using the rules given in Section II, we see that an infinity of separable processes can be constructed. For instance, we could take any number of the examples given in Section VI, subtract their means, and multiply them together. Thus, for example, the product of a zero-mean Gaussian process and a sine wave is a separable process. Also, by adding together processes with identical spectra, we can construct more separable processes. Thus the separable class is seen to contain many widely different members, and no over-all characteristic is apparent.

Extensions to nonstationary examples are apparent; the first seven examples are all separable, even if nonstationary.

VII. GENERALIZATIONS OF THE SEPARABLE CLASS OF RANDOM PROCESSES

7.1 SEPARABILITY OF VARIOUS DEGREES

Thus far, we have been considering processes for which

$$g(x_2, \tau) = \int (x_1 - \mu)p(x_1, x_2; \tau)\, dx_1 = g_1(x_2)\rho(\tau)$$

Let us now define a process as being separable of degree n if

$$g(x_2, \tau) = \int (x_1 - \mu)p(x_1, x_2; \tau)\, dx_1 = \sum_{k=1}^{n} h_k(x_2)P_k(\tau)$$

Examples of processes in this class are easily created by adding together n independent separable processes. Let

$$y(t) = \sum_{k=1}^{n} x_k(t)$$

Then

$$f_y(\xi_1, \xi_2; \tau) = \prod_{k=1}^{n} f_k(\xi_1, \xi_2; \tau)$$

and

$$\begin{aligned} G_y(\xi, \tau) &= G_1(\xi, \tau)f_2(\xi) \dots f_n(\xi) \\ &\quad + f_1(\xi)G_2(\xi, \tau) \dots f_n(\xi) \\ &\quad + \dots \\ &\quad + f_1(\xi)f_2(\xi) \dots G_n(\xi, \tau) \\ &= \rho_1(\tau)f_1'(\xi)f_2(\xi) \dots f_n(\xi) \\ &\quad + \dots \\ &\quad + f_1(\xi)f_2(\xi) \dots f_n'(\xi)\rho_n(\tau) \\ &= \sum_{k=1}^{n} H_k(\xi)\rho_k(\tau) \end{aligned}$$

Therefore

$$g_y(y, \tau) = \frac{1}{j2\pi}\int e^{-j\xi(y-\mu_y)} G_y(\xi, \tau)\, d\xi = \sum_{k=1}^{n} h_k(y)\rho_k(\tau)$$

Since each h_k, or H_k, depends on different properties, the sum in $g_y(y, \tau)$ cannot, in general, be reduced, and we have a process separable of degree n. Thus, an amplitude-modulated process (section 6.6) is a separable process of degree two.

We note also that for $z(t) = x(t) + y(t)$, where

$$g_x(z, \tau) = \sum_{k=1}^{m} h_k^{(x)}(z)P_k^{(x)}(\tau)$$

and

$$g_y(z, \tau) = \sum_{\ell=1}^{n} h_\ell^{(y)}(z)P_\ell^{(y)}(\tau)$$

we have

$$G_z(\xi, \tau) = G_x(\xi, \tau)f_y(\xi) + f_x(\xi)G_y(\xi, \tau)$$

$$= \sum_{k=1}^{m} H_k^{(x)}(\xi)P_k^{(x)}(\tau)f_y(\xi) + \sum_{\ell=1}^{n} f_x(\xi)H_\ell^{(y)}(\xi)P_\ell^{(y)}(\tau)$$

$$= \sum_{k=1}^{m+n} H_k(\xi)P_k(\tau)$$

Then

$$g_z(z, \tau) = \sum_{k=1}^{m+n} h_k(z)P_k(\tau)$$

Thus the sum of two processes, one separable of degree m, the other separable of degree n, yields, in general, a process separable of degree m + n. If special conditions are satisfied, the degree is less than m + n. Witness the requirements (in Section II) for m = n = 1.

By an approach analogous to that used in Section II, it is seen that the product of two independent separable processes yields, in general, a process separable of degree three. If the means are zero, we obtain the familiar separable process of degree one.

If a process is separable of degree n,

$$g(x, \tau) = \sum_{k=1}^{n} h_k(x)P_k(\tau)$$

then the input-output crosscovariance function of a nonlinear device, given in Eq. 19, is

$$\Phi_f(\tau) = \int f(x)g(x, \tau)\, dx$$

$$= \sum_{k=1}^{n} C_{f_k} P_k(\tau)$$

where

$$C_{f_k} = \int f(x) h_k(x)\, dx$$

Thus, the input-output crosscovariance function for a separable process of degree n must be a linear combination of the n factors of the g-function. Also, the input autocovariance function is

$$\phi(\tau) = \int x g(x, \tau)\, dx$$

$$= \sum_{k=1}^{n} d_k P_k(\tau)$$

where

$$d_k = \int x h_k(x)\, dx$$

So there is no obvious relation between $\Phi_f(\tau)$ and $\phi(\tau)$ for processes separable of degree $n \geq 2$, except that each must be a linear combination of given functions.

There also exist processes whose degree is infinite; that is, there are processes for which the g-function is expressible in the form of a sum of separable functions only if the sum is an infinite sum. For example, let

$$y(t) = e^{x(t)}$$

where x(t) is a Gaussian process with zero mean. Then

$$G_y(\xi, \tau) = \overline{j[y(t) - \overline{y(t)}]\, e^{j\xi[y(t+\tau) - \overline{y(t+\tau)}]}}$$

$$= j e^{-j\xi \overline{y(t)}} \left[\overline{y(t)\, e^{j\xi y(t+\tau)}} - \overline{y(t)}\; \overline{e^{j\xi y(t+\tau)}} \right]$$

Now, if we show that

$$\overline{y(t)\, e^{j\xi y(t+\tau)}}$$

can only be expressed as an infinite sum of separable functions, then $G_y(\xi, \tau)$ must be also. Now

$$\overline{y(t)\, e^{j\xi y(t+\tau)}} = \overline{e^{x(t)} e^{j\xi e^{x(t+\tau)}}}$$

$$= \iint e^{x_1} e^{j\xi e^{x_2}} p(x_1, x_2; \tau)\, dx_1\, dx_2$$

$$= \sum_{n=0}^{\infty} \frac{(j\xi)^n}{n!} f_n(\tau) \tag{51}$$

where

$$f_n(\tau) = \iint e^{x_1} e^{nx_2} p(x_1, x_2; \tau)\, dx_1\, dx_2$$

$$= \exp\left[\frac{\sigma^2}{2}(1 + n^2 + 2n\rho(\tau))\right]$$

The set of powers $\{\xi^n\}$ is linearly independent, as are the functions $f_n(\tau)$. Then the sum in Eq. 51 could never be made less than an infinite sum, and the process is separable of infinite degree.

A diagram of the classes of possible second-order probability density functions is shown in Fig. 2.

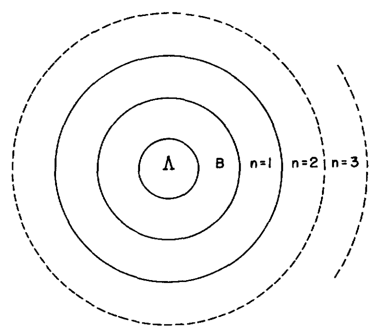

Fig. 2. Classes of separable processes.

The area included in each circle represents a particular class of processes, and includes all of the other circles (classes) of smaller radii (generality). There is an infinity of such circles, with a process existing in each and every circle. The smallest circle, Λ, is Barrett and Lampard's class (8), B is Brown's class (9), and the circle of index n is a class separable of degree n. Note that the class $n = 1$ is the class that satisfies the invariance property. The classes for $n \geq 2$ do not.

7.2 SEPARABILITY OF VARIOUS ORDERS

Another way to generalize the useful results of the separable class is to consider g_3 (and, possibly, higher orders of g-functions):

$$g_3(x_2, x_3; \tau_1, \tau_2) = \int (x_1 - \mu) p(x_1; x_2, \tau_1; x_3, \tau_2)\, dx_1 \tag{52}$$

where $p(x_1; x_2, \tau_1; x_3, \tau_2)$ is the third-order probability density function of the process. Here, also, there are cases that are separable of various degrees. Thus, in general, we would have to investigate processes separable of order m and degree n.

Some questions of separability can be answered, in fact, only by knowing the higher-order properties described in Eq. 52. For example, consider the following processes, all of which are indeterminate, as far as separability is concerned:

(a) sum or product of dependent processes;

(b) sum of process and process delayed;

(c) separable process through linear network; and

(d) inverse of separable process.

Thus to answer questions of separability for processes a and b of this list, we need some results on fourth-order statistics; process c requires knowledge of all orders of input probability density functions; process d is merely a matter of coincidence that depends on the particular second-order probability density function.

Acknowledgment

The author is grateful to Professor Yuk-Wing Lee who supervised the thesis and gave many constructive suggestions. His unceasing friendly interest in the thesis topic made the project very enjoyable.

References

1. N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications (The Technology Press, M. I. T., and John Wiley and Sons, Inc., New York, 1949).
2. C. E. Shannon, The Mathematical Theory of Communication (University of Illinois Press, Urbana, Illinois, 1949).
3. M. C. Wang and G. E. Uhlenbeck, On the theory of the Brownian motion II, Revs. Modern Phys. 17, 323-342 (1945).
4. Y. W. Lee, Application of statistical methods to communication problems, Technical Report 181, Research Laboratory of Electronics, M. I. T., Sept. 1, 1950.
5. P. M. Woodward, Probability and Information Theory (Pergamon Press, Ltd., London, 1953).
6. J. J. Bussgang, Crosscorrelation functions of amplitude-distorted Gaussian signals, Technical Report 216, Research Laboratory of Electronics, M. I. T., March 26, 1952.
7. R. D. Luce, Quarterly Progress Report, Research Laboratory of Electronics, M. I. T., April 15, 1953, p. 37.
8. J. F. Barrett and D. G. Lampard, An expansion for some second-order probability distributions and its applications to noise problems, Trans. IRE, vol. PGIT-1, no. 1, pp. 10-15 (March 1955).
9. J. L. Brown, On a crosscorrelation property for stationary random processes, Trans. IRE, vol. PGIT-3, no. 1, pp. 28-31 (March 1957).
10. R. Courant and D. Hilbert, Methods of Mathematical Physics (Interscience Publishers, Inc., 1st English edition, 1943), Vol. 1, pp. 52, 110.
11. R. C. Booton, Jr., Nonlinear control systems with statistical inputs, Report 61, Dynamic Analysis and Control Laboratory, M. I. T., March 1, 1952.
12. R. C. Booton, Jr., M. V. Mathews, and W. W. Seifert, Nonlinear servomechanisms with random inputs, Report 70, Dynamic Analysis and Control Laboratory, M. I. T., Aug. 20, 1953.
13. R. C. Booton, Jr., The analysis of nonlinear control systems with random inputs, Proc. Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, April 23-24, 1953, Vol. II, pp. 369-391.
14. L. A. Zadeh, A contribution to the theory of nonlinear systems, J. Franklin Inst. 255, 387-408 (May 1953).
15. L. A. Zadeh, Optimum nonlinear filters, J. Appl. Phys. 24, 396-404 (1953).
16. J. L. Doob, Brownian motion and stochastic equations, Ann. Math. 43, 351 (1942).