VII. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee
Prof. A. G. Bose
D. A. Chesler
W. M. Cowan, Jr.
D. A. George
D. L. Haas
A. D. Hause
I. M. Jacobs
K. L. Jordan, Jr.
A. H. Nuttall
T. V. D. Pampoucas
M. Schetzen
D. W. Tufts
C. E. Wernlein
G. D. Zames

Foreword

Continuing our report of Dr. Norbert Wiener's lectures on "Nonlinear Problems in Random Theory," which was started in the Quarterly Progress Report of April 15, 1958, we present:

Lecture 3. Orthogonal Functionals
Lecture 4. Orthogonal Functionals and Autocorrelation Functions
Lecture 5. Application to Frequency Modulation Problems – I
Lecture 6. Application to Frequency Modulation Problems – II

Y. W. Lee, A. G. Bose

A. NONLINEAR PROBLEMS IN RANDOM THEORY* (continued) by Norbert Wiener

Lecture 3

Orthogonal Functionals

I am going to discuss the hierarchy of orthogonal functionals. Suppose that we have a second-degree function $K(\tau_1, \tau_2)$. In order for us to have sufficient hypotheses to work with (although we shall remove some of these hypotheses later), let us take a finite sum for $K(\tau_1, \tau_2)$

$$K(\tau_1, \tau_2) = \sum_n a_n \phi_n(\tau_1) \phi_n(\tau_2) \tag{3.1}$$

Thus we avoid all troubles of rigor. Now, I shall be working with an expression of the form

$$\iint K(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) \tag{3.2}$$

The function $K(\tau_1, \tau_2)$ is assumed to be symmetrical, by the way, although there is no restriction in considering a function of this sort symmetrical. If it is not symmetrical, I interchange the τ_1, τ_2, add it to itself, and divide by 2. This does not change the final quantity in expression 3.2. Similarly, given a functional such as

$$\iiint K(\tau_1, \tau_2, \tau_3)\, dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a) \tag{3.3}$$

I can symmetrize it, and so on.

Now notice the following: The sum of any two expressions of the form of expression 3.2 is an expression of the same sort, and similarly, the sum of any two expressions of the form of expression 3.3 is an expression of the same sort. The functionals in expressions 3.2 and 3.3 are homogeneous polynomial functionals of the Brownian motion. The homogeneous functional of zero order is K_o. The homogeneous functional of the first order is given by

$$\int K(\tau)\, dx(\tau, \alpha) \tag{3.4}$$

The homogeneous functional of the second order is given by expression 3.2, and so on. I can now get nonhomogeneous functionals of the Brownian motion of any degree I desire. For example, I shall call

$$\int K_1(\tau)\, dx(\tau, \alpha) + K_o \tag{3.5}$$

a nonhomogeneous functional of the first degree. These functionals will be functions of α belonging to L^2. We have no trouble in proving that, if the kernels themselves are functions in L^2.

I now want to do the following things: First, to take the constant and normalize it; then to take a first-degree homogeneous functional plus a constant, make it orthogonal to all constants, and normalize. Next, I shall take the homogeneous expression of the second degree plus a homogeneous expression of the first degree plus the zeroth degree. I make that orthogonal to all constants and functionals of the first degree, and so on. In this way, we get a hierarchy of functionals of different degrees, each of them orthogonal to all functionals of lower degree. This is important because it enables us to express a given function of α in terms of orthogonal functionals of different degrees.

Let us start with the zeroth degree, K_o, which is a constant. The mean of the square of K_o is K_o^2, and its absolute value is 1 (I am dealing here with reals); that is, K_o is ± 1. We then have the zeroth-degree normalized functional.

Now let us consider the first-degree expression as given in expression 3.5. I want this to be orthogonal to all zero-degree functionals. But notice that changing Brownian motion into its negative does not change the distribution of Brownian motion. Therefore, it follows that if we multiply the first term of expression 3.5 by a constant and average, the average will be zero. Then multiplying expression 3.5 by a constant C and averaging, we get CK_o. It must be zero if expression 3.5 is to be orthogonal to all constants, which is the case only if $K_o = 0$. Therefore, all first-degree homogeneous functionals are orthogonal to all zero-degree homogeneous functionals. The constant term will have to be zero to make expression 3.5 orthogonal. We now normalize the first-degree orthogonal functional:

$$\int K_1(\tau)\, dx(\tau, a) \tag{3.6}$$

Now

$$\int_0^1 da \left[\int K_1(\tau)\, dx(\tau, a)\right]^2 = \int K_1^2(\tau)\, d\tau \tag{3.7}$$

Thus we have our category of (homogeneous) first-degree functionals, orthogonal with respect to a to all homogeneous zero-degree functionals. These will be represented by

$$\int K_1(\tau)\, dx(\tau, a) \tag{3.8}$$

where

$$\int K_1^2(\tau)\, d\tau = 1 \tag{3.9}$$

So now we have two hierarchies: the zero-degree functionals that are normalized; and the first-degree functionals that are orthogonal to all zero-degree functionals, and are normalized.

The computations now become a little more complicated. I consider a second-degree functional like

$$\iint K_2(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) + \int K_1(\tau)\, dx(\tau, a) + K_o \tag{3.10}$$

where K_2 is symmetrical. I am assuming that K_2 can be represented as the sum of products, as in Eq. 3.1. Now I want expression 3.10 to be orthogonal to every constant, so it is enough to say that it is orthogonal to 1. To say that expression 3.10 is orthogonal to 1 is to say that the average of expression 3.10 multiplied by 1 is 0. This yields

$$\int K_2(\tau, \tau)\, d\tau + K_o = 0 \tag{3.11}$$

We also want to make expression 3.10 orthogonal to any expression

$$\int C(\tau)\, dx(\tau, a) \tag{3.12}$$

If we do that, the first term in the product of expressions 3.10 and 3.12 is third degree and has a zero average. The second term is second degree; the last term is first degree and has a zero average. We then get

$$\int K_1(\tau)\, C(\tau)\, d\tau = 0 \tag{3.13}$$

That means, since Eq. 3.13 is true for any $C(\tau)$, that $K_1(\tau) = 0$. Also, from Eq. 3.11, we have

$$K_o = -\int K_2(\tau, \tau)\, d\tau \tag{3.14}$$

Therefore, the expression that is orthogonal to every zero-degree and first-degree expression is

$$\iint K_2(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) - \int K_2(\tau, \tau)\, d\tau \tag{3.15}$$

We have orthogonalized expression 3.15 to every zero- and first-degree expression, and we shall now normalize it. I square expression 3.15 and integrate with respect to a from 0 to 1:

$$\int_0^1 da \left[\iint K_2(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) - \int K_2(\tau, \tau)\, d\tau\right]^2 \tag{3.16}$$

This becomes

$$\int_0^1 da \left[\iiiint K_2(\tau_1, \tau_2)\, K_3(\tau_3, \tau_4)\, dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a)\, dx(\tau_4, a)\right.$$
$$\left. - 2\int K_2(\tau, \tau)\, d\tau \iint K_2(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) + \left(\int K_2(\tau, \tau)\, d\tau\right)^2\right] \tag{3.17}$$

Now remember our rule: We identify the variables by pairs and integrate. There are three ways of identifying the variables τ_1, τ_2, τ_3, and τ_4 in the first term of expression 3.17 by pairs: τ_1 and τ_2, τ_3 and τ_4; τ_1 and τ_3, τ_2 and τ_4; τ_1 and τ_4, τ_2 and τ_3. Remember that K_2 can be, and is chosen, symmetrical. From expression 3.17 we get

$$\left[\int K_2(\tau, \tau)\, d\tau\right]^2 + 2\iint K_2^2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2$$
$$- 2\left[\int K_2(\tau, \tau)\, d\tau\right]^2 + \left[\int K_2(\tau, \tau)\, d\tau\right]^2 \tag{3.18}$$

where the first two terms result from identification of the variables $\tau_1, \tau_2, \tau_3, \tau_4$ in various combinations. Summing, we get

$$2 \iint K_2^2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 \tag{3.19}$$

So now we have the second category of orthogonal functionals. In this category we have

$$\iint K_2(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha) - \int K_2(\tau, \tau)\, d\tau \tag{3.20}$$

where

$$2 \iint K_2^2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 = 1 \tag{3.21}$$

I shall construct the third category in order to give you the feel of this. Then I shall go over to the general theory. We take

$$\begin{aligned} &\iiint K_3(\tau_1, \tau_2, \tau_3)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\, dx(\tau_3, \alpha) \\ &\quad + \iint K_2(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha) \\ &\quad + \int K_1(\tau)\, dx(\tau, \alpha) + K_0 \end{aligned} \tag{3.22}$$

where K_3 is symmetrical. This is the third-degree nonhomogeneous functional of $x(t, \alpha)$. This is to be orthogonalized to a constant, to a first-degree functional, and to a second-degree functional. First, consider the constant. The average of the first term of expression 3.22 is zero because it is third degree; the second term of expression 3.22 yields

$$\int K_2(\tau, \tau)\, d\tau \tag{3.23}$$

The average of the third term is zero, and the last term gives K_0. Thus we have

$$\int K_2(\tau, \tau)\, d\tau + K_0 = 0 \tag{3.24}$$

Also, expression 3.22 is supposed to be orthogonal to any expression

$$\int C(\tau)\, dx(\tau, \alpha) \tag{3.25}$$

So, we multiply expression 3.22 by expression 3.25 and average. Since K_3 is symmetrical, all of the ways of dividing τ_1, τ_2, τ_3 in pairs are alike, and we obtain

$$3\iint K_3(\tau_1, \tau_1, \tau)\, C(\tau)\, d\tau\, d\tau_1 + \int K_1(\tau)\, C(\tau)\, d\tau = 0 \tag{3.26}$$

for all $C(\tau)$. And since $C(\tau)$ is arbitrary, the function that is orthogonal to every function $C(\tau)$ is zero. Therefore, we obtain

$$3 \int K_3(\tau_1, \tau_1, \tau)\, d\tau_1 + K_1(\tau) = 0, \text{ for all } \tau \tag{3.27}$$

That is the necessary and sufficient condition for orthogonality with linear functionals. Next, we multiply expression 3.22 by

$$\iint C(\sigma_1, \sigma_2)\, dx(\sigma_1, a)\, dx(\sigma_2, a) \tag{3.28}$$

where $C(\sigma_1, \sigma_2)$ is symmetric, and then average. That gives us a fifth-degree expression. The first and third terms of the fifth-degree expression have zero averages. We then get

$$\int C(\sigma, \sigma)\, d\sigma \int K_2(\tau, \tau)\, d\tau + 2 \iint C(\tau_1, \tau_2)\, K_2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2$$

$$+ K_0 \int C(\sigma, \sigma)\, d\sigma = 0 \tag{3.29}$$

where the first term results from identifying σ's and τ's among themselves, and the second term results from identifying σ's with τ's. Then if we use Eq. 3.24, we see that the first term plus the third term in Eq. 3.29 is zero. Now, whatever $C(\sigma_1, \sigma_2)$ is, Eq. 3.29 must be zero. Also, any symmetric function orthogonal to all symmetric functions is identically zero. Therefore, any K_2 is identically zero, and from Eq. 3.27 we have

$$K_1(\tau) = -3 \int K_3(\tau_1, \tau_1, \tau)\, d\tau_1, \text{ for all } \tau \tag{3.30}$$

Using Eq. 3.30 in expression 3.22, we find the third-degree functional to be

$$\iiint K_3(\tau_1, \tau_2, \tau_3)\, dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a)$$

$$- 3 \iint K_3(\tau_1, \tau_1, \tau)\, d\tau_1\, dx(\tau, a) \tag{3.31}$$

This has been orthogonalized to every constant, linear, and quadratic functional, but it has not yet been normalized. So we want

$$1 = \int_0^1 da \left[\iiint K_3(\tau_1, \tau_2, \tau_3)\, dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a) - 3 \iint K_3(\tau_1, \tau_1, \tau)\, d\tau_1\, dx(\tau, a)\right]^2 = \int_0^1 da \left[\iiint\!\!\iiint K_3(\tau_1, \tau_2, \tau_3)\, K_3(\tau_4, \tau_5, \tau_6)\right.$$
$$\times dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a)\, dx(\tau_4, a)\, dx(\tau_5, a)\, dx(\tau_6, a) - 6 \iiint\!\!\iint K_3(\tau_1, \tau_2, \tau_3)\, K_3(\sigma_1, \sigma_1, \tau)\, d\sigma_1\, dx(\tau, a)\, dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a)$$
$$\left. + 9 \iiiint K_3(\tau_1, \tau_1, \tau_2)\, K_3(\tau_3, \tau_3, \tau_4)\, d\tau_1\, d\tau_3\, dx(\tau_2, a)\, dx(\tau_4, a)\right] \tag{3.32}$$

The first term on the right-hand side of Eq. 3.32 is a sixfold integral, and we can identify the variables in a number of different ways. One of the things that we can do is to identify every τ in the first parentheses with some τ in the second parentheses. There are three ways of dealing with the first bracket; and after that, there are only two ways of picking the second and one way for the third. So we get

$$3! \iiint K_3^2(\tau_1, \tau_2, \tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 \tag{3.33}$$

for the first combination. The second combination is obtained by identifying one τ in the first parentheses of Eq. 3.32 with another τ in the first parentheses. This means that we have one τ left over which will be identified with one τ in the second parentheses, and there will be two τ's left over from that. Then we have

$$\iiint K_3(\tau_1, \tau_1, \tau)\, K_3(\tau_2, \tau_2, \tau)\, d\tau_1\, d\tau_2\, d\tau \tag{3.34}$$

as a typical term. How many ways are there of obtaining that term? The single term in the first parentheses of Eq. 3.32 can be picked out in three ways. The single term in the other parentheses can also be picked out in three ways. So we have nine ways of doing it. Thus, the first term of Eq. 3.32 yields 9 times expression 3.34 plus expression 3.33. Now we take

$$-6 \int_0^1 da \iiint\!\!\iint K_3(\tau_1, \tau_2, \tau_3)\, K_3(\sigma_1, \sigma_1, \tau)\, d\sigma_1\, dx(\tau, a)$$
$$\times dx(\tau_1, a)\, dx(\tau_2, a)\, dx(\tau_3, a) \tag{3.35}$$

which is the second term result in Eq. 3.32. Now, τ will have to be identified with one of the τ_1, τ_2, τ_3, for there is no other way to do it. That means that the two remaining τ's are to be identified with one another, and we get exactly the same thing as expression 3.34 except for a scale factor of -6×3, or -18. There is left the last term of Eq. 3.32. There we identify τ_2 and τ_4 and obtain

$$9 \int \left[\int K_3(\tau_1, \tau_1, \tau)\, d\tau_1\right]^2 d\tau \tag{3.36}$$

Equation 3.32 becomes

$$3! \iiint K_3^2(\tau_1, \tau_2, \tau_3)\, d\tau_1\, d\tau_2\, d\tau_3$$

$$+ 9 \iiint K_3(\tau_1, \tau_1, \tau)\, K_3(\tau_2, \tau_2, \tau)\, d\tau_1\, d\tau_2\, d\tau$$

$$- 18 \iiint K_3(\tau_1, \tau_1, \tau)\, K_3(\tau_2, \tau_2, \tau)\, d\tau_1\, d\tau_2\, d\tau$$

$$+ 9 \int \left[\int K_3(\tau_1, \tau_1, \tau)\, d\tau_1\right]^2 d\tau \tag{3.37}$$

The last three integrals in expression 3.37 are the same – just check out the variables. So we get

$$\iiint K_3(\tau_1, \tau_2, \tau_3)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\, dx(\tau_3, \alpha)$$

$$- 3 \iint K_3(\tau_1, \tau_1, \tau)\, d\tau_1\, dx(\tau, \alpha) \tag{3.38}$$

where

$$3! \iiint K_3^2(\tau_1, \tau_2, \tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 = 1 \tag{3.39}$$

I have given you the first three orthogonal functionals. Actually, I could construct the fourth. The fourth-degree functional will contain the fourth-order, second-order, and zeroth-order terms, and so on. In any case, the leading term (the term of highest order) will determine the other terms; and the integral of the square of the orthogonal functional, if it is not normalized, will be the integral of the square of the highest symmetric term, with respect to the three variables multiplied by 3!, or, more generally, with respect to the n variables multiplied by n!

Consider, once more, the second-degree-functional case (the third-order case will be similar). Notice that this function of α (expression 3.15) is related to a normalized symmetric function of τ_1, τ_2 in such a way that if we divide K_2 in expression 3.15 by $\sqrt{2}$, the integral of the square, expression 3.16, will become the integral of the square, Eq. 3.21, and will be 1. I now employ the Riesz-Fischer theorem. Suppose that we have a sequence of symmetrical second-order functionals $\{K_{2n}(\tau_1, \tau_2)\}$. Then if

$$\lim_{m,n\to\infty} \iint \left[K_{2m}(\tau_1, \tau_2) - K_{2n}(\tau_1, \tau_2)\right]^2 d\tau_1\, d\tau_2 = 0 \tag{3.40}$$

the functions K_{2n} converge in the mean to a function $K_2(\tau_1, \tau_2)$. It follows that the expressions of the form of expression 3.2 converge in the mean in a to a limit that is independent of the sequence chosen. This means that although I have started with rather special functions $K(\tau_1, \tau_2)$, as in Eq. 3.1, which are finite sums of products of functions in L^2, convergence in the mean of these functions will define an operator on a that will be the limit in the mean of the approximation. I shall define this operator as

$$G_2(K_2, a) = \iint K_2(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) - \int K_2(\tau, \tau)\, d\tau \tag{3.41}$$

Notice, then, I shall define a functional $G_2(K_2, a)$ for every function K_2 which is symmetrical, or can be made symmetrical, and which is L^2. I shall have to be a little careful here. It does not follow that I can immediately use the representation in Eq. 3.41, for the following reason: The second term of Eq. 3.41 may not exist even if K_2 is in the class L^2 in the two variables together. Therefore

$$\int K_2(\tau, \tau)\, d\tau < \infty \tag{3.42}$$

is an added requirement. However, we can still define $G_2(K_2, a)$ as a well-defined function of a belonging to L^2 when K_2 is any symmetrical function of τ_1, τ_2 belonging to L^2. Similarly, we can define $G_3(K_3, a)$, $G_1(K_1, a)$, and $G_o(K_o, a)$; the last does not actually depend on a.

I am now going ahead of the game and assume that I proved this for general n, which I shall, in fact, do later. Thus, I can define a hierarchy of functionals of a. Furthermore, it is easy to see that all of the functionals at each level are a closed set in themselves; that is, the sum of any two functionals at each level will be a limit in the mean of functionals at this level because of the convergence in the mean property. In other words, we obtain a series of classes of functionals of a which are L^2. Every functional in each class is orthogonal to every functional in every other class. If in each class we are given a definite $K_o, K_1, K_2, \ldots$, we have a hierarchy of functionals of a. Furthermore, I say that it can be proved that this is a closed set of functionals, that every function of a belonging to L^2 can be approximated by a polynomial functional of this set, and that, since the G's are orthogonal, these approximations represent the projections of the function of a on successive classes or spaces — spaces of all functions of a. In other words, we can represent the function of a completely in this way. (We shall go over that at the next lecture.)

Then, given any function $F(a)$ belonging to L^2, $F(a)$ can be represented in a unique way as

$$F(a) = \underset{N\to\infty}{\text{l.i.m.}} \left[G_o(K_o, a) + G_1(K_1, a) + \ldots + G_N(K_N, a)\right] \tag{3.43}$$

Furthermore, previously determined terms are not changed by taking more terms in the approximation to $F(\alpha)$. This is just like ordinary orthogonal development, except that here we have an orthogonal development in whole classes of functionals – something that will give us a canonical development of any function of α that belongs to L^2. We are going to find that this canonical development in a functional of a random function is extremely useful, and we shall go into that in more detail next time.

Now, we actually have more than the mere knowledge that we can get a development of the sort that we obtained in Eq. 3.43. Indeed, we have a technique by which, given a function of α, we can pick out the respective components. This means picking out $K_0, K_1, K_2, \ldots$. This is the next thing that I am going to talk about.

I want to call your attention to the fact that this development is analogous to the Fourier integral theory, which is a development of a particular sort in which we get the coefficients of the "expansion" of the function by an integration process. Here we have a development of another sort which depends upon $x(t, \alpha)$. But $x(t, \alpha)$ is a well-defined function. We have already defined it. The present development is different from the ordinary Fourier integral development in the following way. We have a denumerable set of functionals, but nevertheless, the general problem is the same: namely, given a function of α, get the development.

Next time I shall show you how to develop explicitly any function of α in L^2 in a canonical development. We shall then use that development for various problems arising in connection with random functions. In particular, we shall use it for the following. Suppose that we have a function not of α alone, but of $T^t\alpha$, where $T^t\alpha$ is obtained by taking $x(\tau, \alpha)$ and changing it to $x(t+\tau, \alpha)$. We have already seen that this is a measure-preserving transformation of α. I am going to show you that when we make this transformation, each term in the development of $F(T^t\alpha)$ comes from one term in the development of $F(\alpha)$, and they do not mix. I shall then be able to compute

$$\int_0^1 d\alpha\, F(T^t\alpha)\, F(\alpha) \tag{3.44}$$

using the fact that orthogonality and normality in α for the G's goes over to orthogonality and normality for each one of the G's, with shifted arguments of the kernels in their respective set of variables. Under these conditions, we shall be able to get the autocorrelation function of nonlinear operators on Brownian motion and their spectra.

References

N. Wiener, Response of a nonlinear device to noise, Report 129, Radiation Laboratory, M.I.T., 1942.

R. H. Cameron and W. T. Martin, The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals, Ann. Math. 48, 385-389 (1947).

N. Wiener, Seminar in nonlinear networks, Research Laboratory of Electronics, M.I.T., Feb. 1949 (unpublished).

N. Wiener, Mathematical problems of communication theory, Summer Session Lecture Notes, M.I.T., 1953-54 (unpublished).

A. G. Bose, A Theory of Nonlinear Systems, Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956.

J. F. Barrett, The use of functionals in the analysis of nonlinear physical systems, Statistical Advisory Unit Report No. 1/57, Ministry of Supply, G.B. (no date given).

K. O. Friedrichs, Integration of functionals, Course Notes, New York University, 1957 (unpublished).

Lecture 4

Orthogonal Functionals and Autocorrelation Functions

Today I am going to discuss the general theory of the development that I gave at the last lecture.

We have

$$F_n(\alpha) = \int \cdots \int K_n(\tau_1, \ldots, \tau_n)\, dx(\tau_1, \alpha) \ldots dx(\tau_n, \alpha)$$
$$+ \int \cdots \int K_{n-1}(\tau_1, \ldots, \tau_{n-1})\, dx(\tau_1, \alpha) \ldots dx(\tau_{n-1}, \alpha) + \ldots + K_0 \tag{4.1}$$

I am assuming, by the way, that K_n has all smooth properties, since I am dealing with a particular case. I shall take K_n as given. The other K's are to be determined so that $F_n(\alpha)$ is orthogonal, as a function of α, to every integral

$$\int \cdots \int Q(\tau_1, \ldots, \tau_m)\, dx(\tau_1, \alpha) \ldots dx(\tau_m, \alpha) \tag{4.2}$$

for which $m < n$.

That being the case, I am interested in obtaining

$$\int_0^1 F_n^2(\alpha)\, d\alpha \tag{4.3}$$

From our assumptions, it follows at once that

$$\int_0^1 F_n^2(\alpha)\, d\alpha = \int_0^1 d\alpha \left[\int \cdots \int K_n(\sigma_1, \ldots, \sigma_n)\, dx(\sigma_1, \alpha) \ldots dx(\sigma_n, \alpha)\right]$$
$$\times \left[\int \cdots \int K_n(\tau_1, \ldots, \tau_n)\, dx(\tau_1, \alpha) \ldots dx(\tau_n, \alpha) + \text{lower terms}\right] \tag{4.4}$$

since all of the lower-order terms will be orthogonal to $F_n(\alpha)$ over $(0, 1)$. In order to perform the integration, we take all of the σ variables and all of the τ variables, make identification of variables by pairs in all possible ways, integrate, and add. Notice that if we identify two σ variables, then we get an expression of lower degree. The integral of such a term multiplied by $F_n(\alpha)$ will be zero. Therefore, the only things that are left are those identifications in which we identify each σ variable with one of the τ variables.

We can only identify each σ variable with a τ variable in the first term of $F_n(\alpha)$. There are $n!$ ways of doing this. Thus we get

$$\int_0^1 F_n^2(a)\, da = n! \int \dots \int K_n^2(\tau_1, \dots, \tau_n)\, d\tau_1 \dots d\tau_n \tag{4.5}$$

In other words, without actually computing the expression, I have obtained the formula for the integral of the square of the n^{th}-degree orthogonalized F_n, at least in the case in which K_n is the finite sum of products of L^2 functions.

I can now use the argument that I gave before. If we have given any limit in the mean (l.i.m.) of expressions of the form of Eq. 4.1, then a necessary and sufficient condition for the limit in the mean to tend to a limit will be that the K_n tend to a limit. Under those conditions, we shall still get the relation of Eq. 4.5.

The general function that I shall get from Eq. 4.1, I shall designate as $G_n(K_n, a)$, where K_n is a symmetric function of n variables. I now extend this by the limit in the mean to all K_n that are L^2 in the different variables. Any $G_n(K_n, a)$ is orthogonal to any $G_m(K_m, a)$ for which $m \neq n$.

The next question is the closure of these functions, but we have really settled that question. All the sets of the G_n are closed, as will be shown. Do you remember how we got the a's? We took $x(t, a)$ and defined it in terms of a permutable binary number. Suppose that we have any function

$$\phi[x(t_1, a)\; x(t_2, a)\; \dots\; x(t_n, a)] \tag{4.6}$$

where the t_n are binary numbers of a certain order. You remember that when I defined the curve that goes through a set of "holes," I defined it in terms of the x's only. In other words, I defined the measure of a in terms of the functions that depended on $x(t, a)$ at a larger and larger binary number of points. It is easy to use that to show closure of functions of a at these points.

However, because of the closure of $x^n \exp(-x^2)$, it is easy to show that ϕ can be approximated by polynomials that pass through the given points. Therefore, any function of a belonging to L^2 can be approximated in the mean by polynomials depending on x's, or differences between x's, at a finite number of points. Since the approximating polynomials can be approximated by orthogonal polynomials, we have at least one approximation of this sort. Therefore the G_n form a closed set if K_n runs through all functions that are symmetrical and L^2 in n variables.

In summary, any function of a can be approximated by a sum of orthogonal G functions. Furthermore, if we take the sum of two G's of n^{th} degree, we get a term of n^{th} degree. If we take the projection of $F(a)$ on all terms of lower degree than n, where n is the degree of the highest term in an approximation to F, this projection does not change as we increase n. The projection of a projection will still be the same projection.

From the preceding discussion we can come to the conclusion that if $F(a)$ belongs to

L^2 we can write, uniquely,

$$F(a) = \underset{N\to\infty}{\mathrm{l.i.m.}} \sum_{\nu=0}^{N} G_\nu(K_\nu, a) \tag{4.7}$$

where the best representation for each choice of N does not change the lower-order G_ν but only introduces new ones.

Now comes the problem of determining the G_ν, which means determining the K's. Suppose that we take

$$G_\mu(Q_\mu, a) \tag{4.8}$$

where G_μ is a given $G_\mu(Q_\mu, a)$. Any expression of that sort will be orthogonal to all of the terms in Eq. 4.7 except the μ^{th}, and we have

$$\int_0^1 da\, F(a)\, G_\mu(Q_\mu, a) = \int_0^1 G_\mu(K_\mu, a)\, G_\mu(Q_\mu, a)\, da \tag{4.9}$$

All other terms will vanish. We have already obtained our formula for this integral. It is

$$\int_0^1 da\, F(a)\, G_\mu(Q_\mu, a) = \mu! \int d\tau_1 \dots \int d\tau_\mu K_\mu(\tau_1, \dots, \tau_\mu)\, Q_\mu(\tau_1, \dots, \tau_\mu) \tag{4.10}$$

In order to show this, I first work with $(Q_\mu + K_\mu)$ in Eq. 4.5, then with $(Q_\mu - K_\mu)$, subtract the results, and divide by 4. Dividing by $\mu!$, I obtain

$$\int d\tau_1 \dots \int d\tau_\mu K_\mu(\tau_1, \dots, \tau_\mu)\, Q_\mu(\tau_1, \dots, \tau_\mu) = \frac{1}{\mu!} \int_0^1 da\, F(a)\, G_\mu(Q_\mu, a) \tag{4.11}$$

Equation 4.11 can be used for many purposes. One of the simplest is the following. Let us consider $Q_\mu(a)$ to be a particular Q_μ, which we define as

$$Q_\mu = \left\{ \begin{array}{l} 1, \text{ for every } \tau_k (1 \le K \le \mu),\ t_k < \tau_k < t_k + \epsilon \\ \\ 0, \text{ elsewhere} \end{array} \right\} \tag{4.12}$$

That is a symmetrical function. If we take this particular Q_μ, Eq. 4.11 reduces to

$$\frac{1}{\mu!} \int_0^1 da\, F(a)\, G_\mu(Q_\mu, a) = \int_{t_1}^{t_1+\epsilon} \dots \int_{t_\mu}^{t_\mu+\epsilon} K_\mu(\tau_1, \dots, \tau_\mu)\, d\tau_1 \dots d\tau_\mu \tag{4.13}$$

From Eq. 4.13 we obtain

$$\frac{1}{\epsilon^\mu \mu!} \int_0^1 da\ F(a)\ G_\mu(Q_\mu, a) \tag{4.14}$$

as the multiple average of K_μ over the ϵ interval. If K_μ is a function belonging to L^2 it can be proved that the average will converge in the mean to K_μ. Therefore, we have a procedure by which we can explicitly determine the K_μ and the expansion of $F(\alpha)$.

To summarize: We have here an expansion in terms of which, given F, we can, by integrating F times the known G's and going to the limit in the mean, determine the different K's and the different terms in the development. This, then, is a formulation analogous to the Fourier integral development in that, given any function belonging to L^2, I can not only say that there is a development like that of Eq. 4.7 but I can determine, explicitly, the coefficients by a limit in the mean of certain integrals. This is the expansion that is important for nonlinear work, just as the Fourier integral is important for linear work.

There are other particular cases that I can work out. One important case is the one in which we have

$$G_\mu(\tau_1, \ldots, \tau_\mu) = \phi_1(\tau_1)\,\phi_2(\tau_2) \ldots \phi_\mu(\tau_\mu) \tag{4.15}$$

where the ϕ's are a normal and orthogonal set of functions.

By the way, it is not necessary for G_μ to be symmetrical. It can always be symmetrized. Some of the ϕ's may be the same, and some of them may be different. I can still get the development of the K's in terms of the ϕ's. I am going to use that later.

Now suppose that we start with

$$F(\alpha) = \text{l.i.m.} \sum_{n=0}^{\infty} G_n(K_n, \alpha) \tag{4.16}$$

and suppose that

$$K_n = K_n(\tau_1, \ldots, \tau_n) \tag{4.17}$$

Let me replace Eq. 4.17 by

$$K_n = K_n(t+\tau_1, \ldots, t+\tau_n) \tag{4.18}$$

I think that it is clear that this replacement is generated by a measure-preserving transformation of α. I think that it is also clear, from the way in which the different terms are formed, that each K_n in Eq. 4.17 will go into the corresponding K_n in Eq. 4.18 under the same transformation, and that the G_n will go into the G_n under the same transformation.

If we designate the transformation of α as T^t then

$$F(T^t\alpha) = \sum_{n=0}^{\infty} G_n(K_n^*, \alpha) \tag{4.19}$$

where

$$K_n^*(\tau_1, \ldots, \tau_n) = K_n(t+\tau_1, \ldots, t+\tau_n) \tag{4.20}$$

First, I shall work with a real case, then say a word or two about the complex case. It follows that

$$\int_0^1 F(T^t\alpha)\, F(\alpha)\, d\alpha = \sum_{n=0}^{\infty} n! \int \cdots \int K_n(\tau_1, \ldots, \tau_n)\, K_n(t+\tau_1, \ldots, t+\tau_n)\, d\tau_1 \ldots d\tau_n \tag{4.21}$$

If the K's are complex, we can separate real and imaginary parts, and the final result will be:

$$\int_0^1 F(T^t\alpha)\, \overline{F(\alpha)}\, d\alpha = \sum_{n=0}^{\infty} n! \int \cdots \int \overline{K_n(\tau_1, \ldots, \tau_n)}\, K_n(t+\tau_1, \ldots, t+\tau_n)\, d\tau_1 \ldots d\tau_n \tag{4.22}$$

There is no problem in proving that.

Equation 4.22 is an autocorrelation obtained by averaging on α, but we now go back to a certain ergodic argument. If we change $x(t, \alpha)$ to $x(t + \sigma, \alpha) - x(\sigma, \alpha)$, we change the Brownian motion in the following way (remember that t runs from $-\infty$ to ∞): Instead of referring the motion to the origin, we refer it to the point P, as shown in Fig. VII-1. We still get a Brownian motion with the same distribution. I think I showed you the other

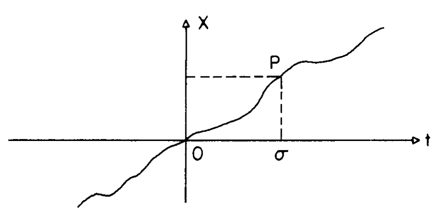

Fig. VII-1.

day that, when we obtain our new α by "cutting down," the measures will be the same. This transformation can be written as

$$x(t+\sigma, \alpha) - x(\sigma, \alpha) = x(t, T^{\sigma}\alpha) \tag{4.23}$$

and is a measure-preserving transformation of α generated by translation. It is such a measure-preserving transformation that is a condition for what is known as metric transitivity.

I shall first state the ergodic theorem in general, then I shall apply it to this case. Suppose that we have a variable α and a measure-preserving transformation $T\alpha$. As a matter of fact, I am going to consider a group $T^t\alpha$ which is such that

$$T^{t_1}\alpha \, T^{t_2}\alpha = T^{t_1+t_2}\alpha \tag{4.24}$$

is a measure-preserving transformation of α. If I assume that I have a function $F(T^t\alpha)$ which belongs to L^2 and is measurable in the pair of variables t and α, then the Birkhoff ergodic theorem says that

$$\lim_{A\to\infty} \frac{1}{A}\int_0^A f(T^t\alpha)\, dt \text{ exists for almost all } \alpha \tag{4.25}$$

It can be proved that the set of values of α for which this integral lies between 0 and ∞ is a measurable set, and it can be proved that this set is invariant under every transformation T^t.

If the transformation T^t has the property of having no invariant measurable set of measure other than 0 or 1, then for almost all α the limit of expression 4.25 is the same. This is called metrical transitivity. It is also known as the ergodic hypothesis. Then, under the ergodic hypothesis, it follows that for any metrically transitive transformation we have

$$\lim_{A\to\infty} \frac{1}{A}\int_0^A f(T^t\alpha)\, dt = C \tag{4.26}$$

for almost all values of α.

What is C? Let us integrate with respect to α:

$$\int_0^1 C\, d\alpha = C = \lim_{A\to\infty} \frac{1}{A}\int_0^A dt \int_0^1 d\alpha\, f(T^t\alpha) \tag{4.27}$$

Then, $f(\alpha)$ is absolutely integrable, and we can interchange orders of integration. Integrating first with respect to t, we get

$$C \equiv \lim_{A\to\infty} \frac{1}{A}\int_0^A f(T^t\alpha)\, dt = \int_0^1 f(\alpha)\, d\alpha \tag{4.28}$$

Hence, if we have a metrically transitive transformation, the time average, for almost all values of α, will be the α average.

From this we deduce that if the metrically transitive hypothesis is fulfilled

$$\int_0^1 F(T^t\alpha)\,\overline{F(\alpha)}\,d\alpha = \lim_{A\to\infty} \frac{1}{A}\int_0^A F(T^t\alpha)\,\overline{F(\alpha)}\,d\alpha \tag{4.29}$$

since

$$F(T^t\alpha)\,\overline{F(\alpha)} \tag{4.30}$$

is obviously a function that will be L^2, being the product of two functions in L^2. Therefore, if we can establish metric transitivity for this particular transformation, we shall have established the time autocorrelation of $F(\alpha)$ for almost all values of α.

Let us go back to the question of metric transitivity. My statement is that the transformation of α which is generated by a translation in time of the Brownian motion is, in fact, metrically transitive; that is, if a set S, of values of α, is invariant under all transformations $T^t\alpha$, its measure is either 1 or 0. I shall sketch the proof.

Consider a set S, of values of α, and assume that S is measurable. There is a theorem for infinite dimensional space which states that if we have a measurable set of values of α, it can be approximated by a finite number of intervals on the α line. A finite number of intervals can, furthermore, be approximated by a finite set of intervals corresponding to the binary subdivisions of the α line. Accordingly, given the set S, there is a set S_f, depending only upon a finite number of points, which is such that

$$m(S\,\overline{S_f} + S_f\,\overline{S}) < \epsilon \tag{4.31}$$

Given any measurable sets, that is the case.

If we take $T^t S_f$, and $|\tau|$ is large enough, then $T^t S_f$ is independent of S_f and S. This means that if we take the logical products $(S\,T^t\,\overline{S_f})$ or $(T^t\,S_f\,\overline{S})$, the measure of the logical product is as close to the product of the measures as we wish. Also, since m(S) is invariant under all transformations T^t, $m(S_f)$ will be nearly invariant. Using these facts in expression 4.31 we obtain, in the limit, since $m(S_f) \approx m(S)$,

$$m(S)\,m(T^t\,\overline{S}) + m(T^t\,S)\,m(\overline{S}) = 2\{m(S) - [m(S)]^2\} < \epsilon \tag{4.32}$$

for any ϵ. Therefore,

$$m^2(S) - m(S) = 0 \tag{4.33}$$

which means that the measure of S is either 0 or 1. Therefore, the translation operator is really a metrically transitive or ergodic transformation.

Summarizing, we can use the Birkhoff ergodic theorem, and we can show that for almost all values of α (and that's all we have to do) the time autocorrelation of $F(\alpha)$ is what we have obtained by our integration formula as the average of the closure.

We are now in a position to go ahead and obtain actual spectra of certain time series that depend nonlinearly on $x(t, \alpha)$. There are two cases of nonlinear spectra that I am going to discuss in later lectures. There are many others that this method will apply to, but I shall discuss these two cases in some detail.

One of these cases is the following:

$$\exp\left(i \int K(t-\tau)\, dx(\tau, \alpha)\right) \tag{4.34}$$

That is essentially what we get, physically, when we have a clock with a locating hand, but the hand is loose and subject to Brownian perturbations. As a matter of fact, expression 4.34 represents a linear Brownian motion of rotation that depends on the parameter x. We are going to take that up at the next lecture, and I shall discuss the spectrum of the motion described by expression 4.34.

This is a very important thing physically; for example, in the sort of problem that Professor Jerrold Zacharias is dealing with — that is, highly perfect clocks. He does not have a perfectly accurate time measurement. The time measurement depends on the clock, as it were. If you have an inaccurate time measurement and a highly perfect clock, how will the time inaccuracies affect the spectrum? We are going to show that the spectrum is as illustrated in Fig. VII-2. Although I have drawn this spectrum centered about zero frequency, I can have it centered at any frequency. The effect of inaccuracies is to take the spectral line and spread it into a band. Actually, the character of this band is not Gaussian, as we shall see.

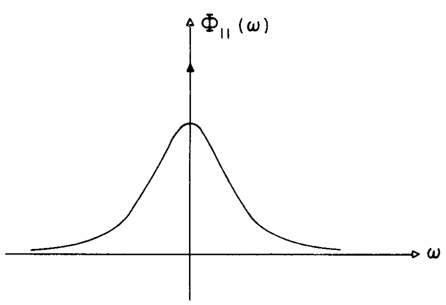

Fig. VII-2.

The next question is similar but one order higher. We have a Brownian motion affecting the speed of a clock in a quadratic way.

$$\exp\left(i \iint K_2(t+\tau_1, t+\tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\right) \tag{4.35}$$

What is the spectrum going to be? I shall tell you in general terms, and we shall compute it at a later date.

Expression 4.35 will give us a line at zero frequency in the spectrum. In addition, there will be sidebands. Their profile will sometimes appear like the one that is shown in Fig. VII-3. However, I am going to show that in certain specified cases the profile

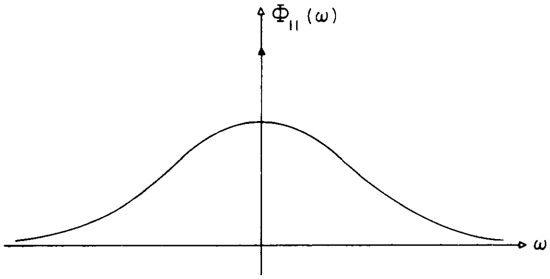

Fig. VII-3.

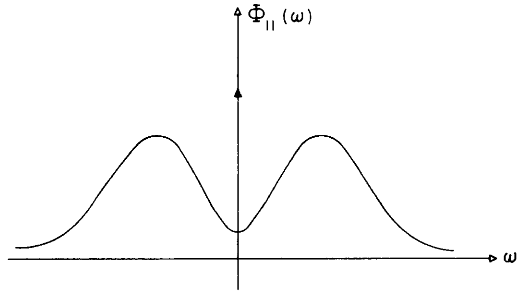

Fig. VII-4.

of the sidebands will be a little different. It will have a dip at the origin, as shown in Fig. VII-4.

The reason that I started this work was that I ran into the case of a spectrum in which there certainly must have been a random element, with sidebands that dip at the center. This came about in my discussion of brain waves. As you know, the electroencephalogram gives a voltage as a function of time. We can get the spectrum of such a function by a method that is familiar to us, and we can get it to a high degree of accuracy. I did this experimentally. I found this dip phenomenon occurring. I tried to think of a possible explanation of it and to get a working model of a system that would exhibit it. The thing that led me into this work was the attempt to get a working model.

The fact is that we do get something of this sort. We are going to begin the evaluation of it at the next lecture.

Lecture 5

Application to Frequency Modulation Problems – I

Today, I want us to consider the function $F_n(\alpha)$ as given by Eq. 5.1.

$$F_n(\alpha) = \int \dots \int K_n(\tau_1, \dots, \tau_n)\, dx(\tau_1, \alpha) \dots dx(\tau_n, \alpha) \tag{5.1}$$

I want to represent $F_n(\alpha)$ in what we call canonical form:

$$F_n(\alpha) = \sum_{k=0}^{n} G_k[L_k(\tau_1, \dots, \tau_k), \alpha] \tag{5.2}$$

where the $\{G_k\}$ are the orthogonal functionals. I want to determine the $L_k(\tau_1, \dots, \tau_k)$ explicitly. In order to do that, I use a certain lemma.

Suppose that I have an expression

$$\sum_{k=0}^{n} \int \cdots \int R_k(\tau_1, \ldots, \tau_k)\, dx(\tau_1, \alpha) \ldots dx(\tau_k, \alpha) \tag{5.3}$$

and I want to multiply this expression by

$$\int \cdots \int K_n(\sigma_1, \ldots, \sigma_n)\, dx(\sigma_1, \alpha) \ldots dx(\sigma_n, \alpha) \tag{5.4}$$

and integrate over α from 0 to 1. If I bring the integration on $K_n(\sigma_1, \ldots, \sigma_n)$ inside the summation, the expression to be evaluated becomes:

$$\int_0^1 d\alpha \sum_{k=0}^{n} \int \cdots \int K_n(\sigma_1, \ldots, \sigma_n)\, R_k(\tau_1, \ldots, \tau_k)$$
$$\times\, dx(\sigma_1, \alpha) \ldots dx(\sigma_n, \alpha)\, dx(\tau_1, \alpha) \ldots dx(\tau_k, \alpha) \tag{5.5}$$

You remember how we integrate this term by term. We have here (k+n) variables in each term. We divide them into pairs in all possible ways, identify the variables of each pair, and integrate. Since n is greater than or equal to k, it is quite clear that when we identify variables by pairs, we must identify variables by pairs among the σ's alone for a certain distance, before we begin identifying variables between the σ's and the τ's, because there are more σ's than τ's. I think it is clear that when I do this, and integrate over only the σ's that are paired with themselves, I shall obtain in each term of the sum an expression,

$$\int \cdots \int K_n(\sigma_1, \ldots, \sigma_k, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_\nu, \lambda_\nu)\, d\lambda_1 \ldots d\lambda_\nu \tag{5.6}$$

where $n = k + 2\nu$. Then I have to begin identifying the σ's with the τ's and carry on the rest of the integrations. It is quite clear that I will get exactly the same results in this two-step integration as I would if I integrated over all the pairings of σ's and τ's in one step. How many times does expression 5.6 enter into each term of the sum? Well, in the first place, how many ways are there for choosing 2ν variables out of n variables? I think it is clear that there are $\binom{n}{2\nu}$ ways, where $\binom{n}{2\nu}$ is the binomial coefficient,

$$\binom{n}{2\nu} = \frac{n!}{(2\nu)!(n-2\nu)!} \tag{5.7}$$

In addition to choosing the 2ν variables I have to identify them in pairs. This gives me $(2\nu - 1)(2\nu - 3) \ldots (1)$ additional terms. Hence, in the integration of expression 5.5, in each term I can replace

$$\int \cdots \int K_n(\sigma_1, \ldots, \sigma_n)\, dx(\sigma_1, \alpha) \ldots dx(\sigma_n, \alpha) \tag{5.8}$$

by

$$\int \cdots \int \binom{n}{2\nu}(2\nu-1)(2\nu-3)\ldots(1) \int \cdots \int K_n(\sigma_1, \ldots, \sigma_k, \lambda_1, \lambda_1, \ldots, \lambda_\nu, \lambda_\nu)$$

$$\times\, d\lambda_1 \ldots d\lambda_\nu\, dx(\sigma_1, \alpha) \ldots dx(\sigma_k, \alpha) \tag{5.9}$$

I can next simplify the coefficient as follows:

$$\binom{n}{2\nu}(2\nu-1)(2\nu-3)\ldots(1) = \frac{n!}{(2\nu)!(n-2\nu)!}\, \frac{(2\nu)!}{(2\nu)(2\nu-2)\ldots(2)}$$

$$= \frac{n!}{(n-2\nu)!}\, \frac{1}{2^\nu\, \nu!} \tag{5.10}$$

From the results of this lemma, it follows at once that if expression 5.3 represents a G-function, we can apply the lemma to the following integration:

$$\int_0^1 d\alpha \int \cdots \int K_n(\tau_1, \ldots, \tau_n)\, dx(\tau_1, \alpha) \ldots dx(\tau_n, \alpha)\, G_k[M_k(\sigma_1, \ldots, \sigma_k), \alpha] \tag{5.11}$$

We shall get exactly what we would get if we replace K_n in expression 5.11 by expression 5.9. I know that this will be

$$K! \int \cdots \int \frac{n!}{(n-2\nu)!\, 2^\nu\, \nu!}\, K_n(\tau_1, \ldots, \tau_k, \lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_\nu, \lambda_\nu)$$

$$\times\, d\lambda_1 \ldots d\lambda_\nu\, M_k(\tau_1, \ldots, \tau_k)\, d\tau_1 \ldots d\tau_k \tag{5.12}$$

Remember, expression 5.11 is equal to expression 5.12 only when k is of the same parity as n. When k is of different parity from n, expression 5.11 is identically zero.

From this, it is easy to come to the conclusion that, for the L_k of Eq. 5.2, the L_k in which k is of different parity from n do not come into the expression. That is, $L_{n-2\nu-1}$ is identically zero in all cases, and the $L_{n-2\nu}$ are

$$L_{n-2\nu}(\tau_1, \ldots, \tau_{n-2\nu}) = \frac{n!}{(n-2\nu)!\, 2^\nu\, \nu!} \int \cdots \int K_n(\tau_1, \ldots, \tau_{n-2\nu}, \sigma_1, \sigma_1, \ldots, \sigma_\nu, \sigma_\nu)$$

$$\times\, d\sigma_1 \ldots d\sigma_\nu \tag{5.13}$$

Equations 5.1, 5.2, and 5.13 can be combined and written as

$$\int \cdots \int K_n(\tau_1, \ldots, \tau_n)\, dx(\tau_1, \alpha) \ldots dx(\tau_n, \alpha) = \sum_{\nu=0}^{[n/2]} \frac{n!}{(n-2\nu)!\, 2^\nu\, \nu!}$$

$$\times G_{n-2\nu}\left[\int \cdots \int K_n(\tau_1, \ldots, \tau_{n-2\nu}, \sigma_1, \sigma_1, \ldots, \sigma_\nu, \sigma_\nu)\, d\sigma_1 \ldots d\sigma_\nu, \alpha\right] \tag{5.14}$$

where $[n/2] = n/2$ if n is even and equals $(n-1)/2$ if n is odd. From this I have obtained the homogeneous polynomial development in terms of the orthogonal functionals that belong to it. These orthogonal functionals are all of the same parity as the K_n, and, for example, in the case of G_ν, are obtained by taking the variables of K_n beyond ν, identifying by pairs, and integrating.

Now, I shall go directly to the use of this development for the study of frequency modulation. I want to study the spectrum of an expression like

$$\exp\left[i \int f(t+\tau)\, dx(\tau, \alpha)\right] \tag{5.15}$$

Notice, of course, that we shall get two terms here, the cosine and the sine, but they can be discussed in terms of expression 5.15. Thus, we have a message that is the response of a linear resonator to a Brownian input, and we are looking for its spectrum. To get the spectrum, we have seen that it is enough to express this sort of thing in terms of the fundamental orthogonal functionals. Now, I am going to make this a bit more general. We shall introduce into the exponent a complex number "a," which, for convenience, will include the factor i; f can now be considered normalized. Hence, expression 5.15 becomes

$$\exp\left[a \int \phi(t+\tau)\, dx(\tau, \alpha)\right] \tag{5.16}$$

where

$$\int \phi^2(\tau)\, d\tau = 1 \tag{5.17}$$

I want the development of expression 5.16 in terms of orthogonal functionals. That is easy. It can be represented as a sum of homogeneous polynomial functionals by simply using the exponential development

$$\exp\left[a \int \phi(\tau)\,dx(\tau,\alpha)\right] = \sum_{n=0}^{\infty} \frac{a^n}{n!} \int \cdots \int \phi(\tau_1) \ldots \phi(\tau_n)\,dx(\tau_1,\alpha) \ldots dx(\tau_n,\alpha) \tag{5.18}$$

where I have let ϕ be a function of τ alone. Later on, I shall reintroduce t. Now, I use Eq. 5.14 for the expression in terms of the orthogonal functionals of each term of the sum of Eq. 5.18, and I obtain

$$\exp\left[a \int \phi(\tau)\,dx(\tau,\alpha)\right] = \sum_{n=0}^{\infty} a^n \sum_{\nu=0}^{[n/2]} \frac{1}{(n-2\nu)!\,2^{\nu}\,\nu!} G_{n-2\nu}[\phi(\tau_1) \ldots \phi(\tau_{n-2\nu}),\alpha] \tag{5.19}$$

where we have paired ϕ's, integrated, and applied Eq. 5.17. Rearranging Eq. 5.19, and letting $\mu = n-2\nu$, we obtain

$$\sum_{\mu=0}^{\infty} \frac{1}{\mu!} G_{\mu}[\phi(\tau_1) \ldots \phi(\tau_{\mu}),\alpha] \sum_{\nu=0}^{\infty} \frac{a^{\mu+2\nu}}{2^{\nu}\,\nu!} \tag{5.20}$$

Equation 5.20 can be simplified a great deal. Summing on ν, it becomes

$$\exp\left[a \int \phi(\tau)\,dx(\tau,\alpha)\right] = \sum_{\mu=0}^{\infty} \frac{a^{\mu}}{\mu!} \exp\left(\frac{a^2}{2}\right) G_{\mu}[\phi(\tau_1) \ldots \phi(\tau_{\mu}),\alpha] \tag{5.21}$$

I have now given, for the FM case, the development in orthogonal polynomial functionals. I am going to do two things with this development. I am going to apply it immediately to the FM spectrum, and I am going to use it as a general tool for handling more complicated cases. First, replace a by ai. We obtain the following function of α:

$$f(\alpha) = \exp\left[ai \int \phi(\tau)\,dx(\tau,\alpha)\right] = \sum_{\mu=0}^{\infty} \frac{(ai)^{\mu}}{\mu!} \exp\left(-\frac{a^2}{2}\right) G_{\mu}[\phi(\tau_1) \ldots \phi(\tau_{\mu}),\alpha] \tag{5.22}$$

If we now replace τ by $t+\tau$, we obtain a function of both t and α:

$$f(t,\alpha) = \exp\left[ai \int \phi(t+\tau)\,dx(\tau,\alpha)\right] = \sum_{\mu=0}^{\infty} \frac{(ai)^{\mu}}{\mu!} \exp\left(-\frac{a^2}{2}\right) G_{\mu}[\phi(t+\tau_1) \ldots \phi(t+\tau_{\mu}),\alpha] \tag{5.23}$$

Next, perform the following integration, using the formula that we have for the integral of the product of two G's.

$$\int_0^1 f(t+s,\alpha)\,\overline{f(s,\alpha)}\,d\alpha = \sum_{\mu=0}^{\infty} \frac{a^{2\mu}}{(\mu!)^2} \exp(-a^2)\,\mu!\left[\int \phi(\tau)\,\phi(t+\tau)\,d\tau\right]^{\mu}$$

$$= \exp(-a^2)\exp\left[a^2 \int \phi(t+\tau)\,\phi(\tau)\,d\tau\right] = \exp\left(a^2\left[\int \phi(t+\tau)\,\phi(\tau)\,d\tau - 1\right]\right) \tag{5.24}$$

where $\overline{f(t,\alpha)}$ is the conjugate of $f(t,\alpha)$. Equation 5.24 is, by the way, the autocorrelation taken on the α-scale, and we have seen by the ergodic theorem, that this is almost always the autocorrelation taken on the t-scale. Hence, Eq. 5.24 gives the autocorrelation explicitly in one of the simple FM cases.

We now want to make a harmonic analysis of this FM case. When we expand Eq. 5.24, we obtain a series of powers of the autocorrelation of ϕ. If we take the Fourier transform, we obtain the power spectrum as a sum of terms which involve the spectrum of ϕ and its respective convolutions. Letting h(t) represent the autocorrelation of ϕ and $H(\omega)$ its power spectrum,

$$h(t) = \int \phi(t+\tau)\,\phi(\tau)\,d\tau \tag{5.25}$$

$$H(\omega) = \frac{1}{(2\pi)^{1/2}} \int h(t)\exp(i\omega t)\,dt \tag{5.26}$$

I take repetitions of $H(\omega), H^1(\omega), \ldots, H^n(\omega)$, where

$$H^1(\omega) = H(\omega) \tag{5.27}$$

and

$$H^n(\omega) = \int_{-\infty}^{\infty} H^{n-1}(\omega+\mu)\,H(\mu)\,d\mu \tag{5.28}$$

Then we obtain the spectrum as a series in the repeated $H(\omega)$'s. This gives us the explicit form for the spectrum of frequency modulation.

I think that the next thing for me to do is to state in ordinary language what I have done. I have treated a message that is the response of a linear resonator to a Brownian input. (The actual messages that I send are not like that, but their distribution will not be too far from that.) Suppose that I send this message by frequency modulation. What would be the spectrum of the transmission? This is the problem that I have solved, and it is a large part of the FM problem. I can go much further than this. Suppose that I have a known message and that to it I add a random error distributed in frequency. I want to determine how the random error in the accuracy of the message will affect the frequency modulation. This problem can be solved by similar methods. I am going to leave this now, but I shall use the results as a tool.

Now I want to discuss a more complicated case:

$$\exp\left[i \iint K(t+\tau_1, t+\tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\right] \tag{5.29}$$

We may call this case "quadratic FM." The first thing to do is to realize that in a large number of cases, $K(\tau_1, \tau_2)$, a symmetric function, can be written:

$$K(\tau_1, \tau_2) = \sum a_n\, \phi_n(\tau_1)\, \phi_n(\tau_2) \tag{5.30}$$

where the ϕ_n are the characteristic functions of the kernel, $K(\tau_1, \tau_2)$, and the a_n are the characteristic numbers. This is the bilinear formula. In these cases, the ϕ_n are real and orthogonal, and the a_n, under certain restrictions which I shall assume at the start, are real and positive. Then, expression 5.29 becomes, at least formally,

$$\exp\left[i \iint K(t+\tau_1, t+\tau_2)\, dx(\tau_1,\alpha)\, dx(\tau_2, \alpha)\right]$$

$$= \Pi \exp\left(i\, a_n \left[\int \phi_n(t+\tau)\, dx(\tau, \alpha)\right]^2\right) \tag{5.31}$$

The first case to handle, then, is simply

$$\exp\left(i\, a_n \left[\int \phi_n(t+\tau)\, dx(\tau, \alpha)\right]^2\right) \tag{5.32}$$

I shall obtain the spectrum by means of the method used in the linear FM case. The next thing is to multiply. When we multiply, there is one very important thing. If we have K's composed of different ϕ's, the product of orthogonal functionals of these K's is an orthogonal functional. Next time I shall go into a general discussion of the spectrum and get phenomena that are more complicated than those that have occurred in the simple FM problem.

Lecture 6

Application to Frequency Modulation Problems – II

In the previous lecture I proved that

$$\exp\left[a \int \phi(t)\, dx(t,\alpha)\right] = \sum_{\nu=0}^{\infty} \frac{a^{\nu}}{\nu!} \exp\left(\frac{a^2}{2}\right) G_{\nu}\left[\phi(\tau_1), \ldots, \phi(\tau_{\nu}), \alpha\right] \tag{6.1}$$

where

$$\int \left|\phi^2(t)\right|\, dt = 1 \tag{6.2}$$

and

$$\mathrm{Re}(a) < 1 \tag{6.3}$$

This time I want to study

$$\exp\left(b\left[\int \phi(t)\, dx(t,\alpha)\right]^2\right) \tag{6.4}$$

which equals

$$\sum_{n=0}^{\infty} \frac{b^n}{n!} \int \phi(\tau_1)\, dx(\tau_1,\alpha) \ldots \int \phi(\tau_{2n})\, dx(\tau_{2n},\alpha) \tag{6.5}$$

Then we see that, when we go to the G-functions, expression 6.5 equals

$$\sum_{n=0}^{\infty} C_{2n}\, G_{2n}\left[\phi(\tau_1), \ldots, \phi(\tau_{2n}), \alpha\right] \tag{6.6}$$

The problem here is the determination of the C_{2n}. This could be done directly, but I am going to use a method of generating functions. Multiplying Eq. 6.1 by expression 6.4 on the left-hand side and by expression 6.6 on the right, and integrating over α, we obtain

$$\int_0^1 d\alpha\, \exp\left(a \int \phi(t)\, dx(t,\alpha)\right) \exp\left(b\left[\int \phi(t)\, dx(t,\alpha)\right]^2\right) = \sum_{n=0}^{\infty} a^{2n} \exp\left(\frac{a^2}{2}\right) C_{2n} \tag{6.7}$$

The right-hand side of Eq. 6.7 follows because

$$\int_0^1 d\alpha\, G_\nu\left[\phi(\tau_1), \ldots, \phi(\tau_\nu), \alpha\right] G_{2n}\left[\phi(\tau_1), \ldots, \phi(\tau_n), \alpha\right] = \begin{cases} \nu! & \text{for } 2n = \nu \\ 0 & \text{otherwise} \end{cases} \tag{6.8}$$

Now let

$$u = \int \phi(t)\, dx(t, \alpha) \tag{6.9}$$

and remember that u has a Gaussian distribution. Also,

$$\int_0^1 u\, d\alpha = 0 \tag{6.10}$$

and

$$\int_0^1 u^2\, d\alpha = 1 \tag{6.11}$$

Therefore

$$\int_0^1 d\alpha\, \exp\left(a \int \phi(t)\, dx(t, \alpha)\right) \exp\left(b \left[\int \phi(t)\, dx(t, \alpha)\right]^2\right) = \int_0^1 d\alpha\, \exp(au)\, \exp(bu^2)$$

$$= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) \exp(au + bu^2)\, du \tag{6.12}$$

where we have made use of our knowledge of the distribution of u, and have obtained an expression that we can integrate. Let

$$v = u(1-2b)^{1/2} \tag{6.13}$$

where we must work with b less than 1/2. This is no difficulty because we want to work with small values of b. Notice that if b is imaginary we can work with large values of b. Now, Eq. 6.12 becomes

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{v^2}{2} + \frac{av}{(1-2b)^{1/2}}\right) \frac{dv}{(1-2b)^{1/2}} \tag{6.14}$$

and completing the square, we obtain

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(v - \frac{a}{(1-2b)^{1/2}}\right)^2}{2}\right) \exp\left(\frac{a^2}{2(1-2b)}\right) \frac{dv}{(1-2b)^{1/2}} \tag{6.15}$$

Now we have

$$\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} \exp\left(-\frac{\left(v-\frac{a}{(1-2b)^{1/2}}\right)^2}{2}\right) dv = 1 \tag{6.16}$$

because, even in a complex plane, moving v up or down as well as right or left does not change that integral. In order to show this, we use the Cauchy theorem. We are integrating an expression along a line, and it makes no difference that we integrate it along a parallel line because the integrand goes down to 0 very rapidly as we go to ∞. Also, there are no singularities inside. So, we use the Cauchy theorem, and the integral is 1. Hence, expression 6.15 equals

$$\frac{1}{(1-2b)^{1/2}} \exp\left(\frac{a^2}{2(1-2b)}\right) \tag{6.17}$$

Substituting this in Eq. 6.7, we obtain

$$\sum_{n=0}^{\infty} a^{2n} C_{2n} = \frac{1}{(1-2b)^{1/2}} \exp\left(-\frac{a^2}{2}\right)\exp\left(\frac{a^2}{2(1-2b)}\right) = \frac{1}{(1-2b)^{1/2}} \exp\left(\frac{ba^2}{1-2b}\right) \tag{6.18}$$

Remember, by the way, that

$$C_{2n+1} = 0 \tag{6.19}$$

Expanding the right-hand side of Eq. 6.18, we have

$$\sum_{n=0}^{\infty} a^{2n} C_{2n} = \frac{1}{(1-2b)^{1/2}} \sum_{n=0}^{\infty} \frac{a^{2n}}{n!}\left(\frac{b}{1-2b}\right)^n \tag{6.20}$$

Equating coefficients of a^{2n}, we obtain

$$C_{2n} = \frac{1}{(1-2b)^{1/2}}\left(\frac{b}{1-2b}\right)^n \frac{1}{n!} \tag{6.21}$$

See how we have saved the trouble of adding up a lot of series by using the generating function.

We can now write

$$\exp\left(b\left[\int \phi(t)\, dx(t, \alpha)\right]^2\right) = \sum_{n=0}^{\infty} \frac{b^n}{n!} \left(\frac{1}{1-2b}\right)^{n+1/2} G_{2n}[\phi(\tau_1), \ldots, \phi(\tau_{2n}), \alpha] \tag{6.22}$$

Next, I shall replace b by ib; there is nothing in the preceding work that prevents this. The imaginary case is even more advantageous than the real because of the fact that the denominator of Eq. 6.21 becomes infinite as b becomes infinite, and now we shall not have to worry about coming to a place where the series will fail to converge. So, we have

$$\exp\left(ib\left[\int \phi(t)\, dx(t, \alpha)\right]^2\right) = \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} \left(\frac{1}{1-2ib}\right)^{n+1/2} G_{2n}[\phi(\tau_1), \ldots, \phi(\tau_{2n}), \alpha] \tag{6.23}$$

Let us consider

$$\exp\left(ib \iint K(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\right) \tag{6.24}$$

where $K(\tau_1, \tau_2)$ is symmetric. For certain cases (cases that represent a generalization of the ordinary Fredholm equation), we shall find

$$K(\tau_1, \tau_2) = \sum_{\nu=0}^{\infty} b_\nu\, \phi_\nu(\tau_1)\, \phi_\nu(\tau_2) \tag{6.25}$$

where the $\phi_\nu(\tau)$ are a set of orthogonal real functions called "characteristic functions," and the b_n are the "characteristic numbers" of the problem. Then expression 6.24 becomes

$$\prod_\nu \sum_{n=0}^{\infty} \frac{(ib)^n}{n!} \left(\frac{1}{1-2ib}\right)^{n+1/2} G_{2n}[\phi_\nu(\tau_1), \ldots, \phi_\nu(\tau_{2n}), \alpha] \tag{6.26}$$

by substitution of Eq. 6.25 in expression 6.24 and use of Eq. 6.23.

Now we want to rearrange this series in a series of orthogonal functionals of different orders, and we go to certain properties of the G's that I have not yet taken up. I want to say that

$$G_n[\phi_j(\tau_1), \ldots, \phi_j(\tau_n), \alpha]\, G_m[\phi_k(\tau_{n+1}), \ldots, \phi_k(\tau_{n+m}), \alpha] \tag{6.27}$$

is a polynomial of degree (n + m) that is orthogonal to all polynomials of lower degree. This is clear because of the independence — any function in one set of variables is independent of any function in another set of variables. Also, we know that an expression of lower degree will have at least one part of lower degree than the corresponding part of expression 6.27. Hence, when we integrate with respect to a, we get 0. This tells us at once how we can rearrange expression 6.26 in orthogonal functionals of lower degrees. Remember that the ϕ_ν are orthogonal to one another. There is an even easier way of doing this.

When we multiply two orthogonal polynomials together we get exactly the same result as though we had multiplied their leading terms together. The leading term will be the leading term, so we consider that there is related to expression 6.26 a function consisting of the leading terms of expression 6.26 – a similar series. We multiply the series of leading terms together, and then we get the terms with which we are going to operate in order to get the G's of the desired expansion. So

$$\prod_\nu \sum_{n=0}^{\infty} \frac{(ib_\nu)^n}{n!} \left(\frac{1}{1-2b_\nu i}\right)^{n+1/2} \left[\int \phi_\nu(\tau)\, dx(\tau, a)\right]^{2n} \tag{6.28}$$

corresponds to the leading terms of expression 6.26. When we multiply out, the terms that we get will be the leading terms of the G's in the desired expansion. Let us explore what this is. Expression 6.28 is

$$\prod_\nu \frac{1}{(1-2ib_\nu)^{1/2}} \exp\left(\left[\int \phi_\nu(\tau)\, dx(\tau, a)\right]^2 \frac{ib_\nu}{1-2b_\nu i}\right) \tag{6.29}$$

which represents the generating operator. I am now going to expand this operator and get a sum of homogeneous operators of the different orders. Then I shall replace these homogeneous operators by the G's of the same leading terms and obtain the desired series. But this leads to a discussion of the functions

$$\sum_\nu \frac{ib_\nu}{1-2ib_\nu} \left[\int \phi_\nu(\tau)\, dx(\tau, a)\right]^2 \tag{6.30}$$

This expression can be written as a homogeneous second-degree expression

$$i \iint R(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a) \tag{6.31}$$

just as we wrote Eq. 6.25. Then expression 6.29 becomes

$$\prod_\nu \left(\frac{1}{(1-2ib_\nu)^{1/2}}\right) \exp\left(i \int\int R(\tau_1, \tau_2)\, dx(\tau_1, a)\, dx(\tau_2, a)\right) \tag{6.32}$$

and this is the function that gives the desired leading coefficients.

The next thing is to investigate

$$\prod_{\nu}\left(\frac{1}{(1-2ib_{\nu})^{1/2}}\right) \tag{6.33}$$

At least formally, this is,

$$\exp\left(-\frac{1}{2}\sum_{\nu} \ln(1-2ib_{\nu})\right) \tag{6.34}$$

To begin with, take

$$|b_{\nu}| < \frac{1}{2} \tag{6.35}$$

Then

$$\frac{1}{2}\sum_{\nu} \ln(1-2ib_{\nu}) = \frac{1}{2}\sum_{\nu}\left[(2ib_{\nu}) + \frac{(2ib_{\nu})^2}{2} + \frac{(2ib_{\nu})^3}{3} + \ldots\right] \tag{6.36}$$

so

$$\frac{1}{2}\sum_{\nu} \ln(1-2ib_{\nu}) = \frac{1}{2}\left[2i\sum_{\nu} b_{\nu} + \frac{(2i)^2}{2}\sum_{\nu} b_{\nu}^2 + \frac{(2i)^3}{3}\sum_{\nu} b_{\nu}^3 + \ldots\right] \tag{6.37}$$

Now, remember that

$$K(\tau_1, \tau_2) = \sum_{\nu} b_{\nu}\, \phi_{\nu}(\tau_1)\, \phi_{\nu}(\tau_2) \tag{6.38}$$

where the ϕ_{ν} are a normal and orthogonal set, and the b_{ν} are real and positive. Then, at least formally, we have

$$\sum_{\nu} b_{\nu} = \int K(\tau, \tau)\, d\tau \tag{6.39}$$

$$\sum_{\nu} b_{\nu}^2 = \iint K^2(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 \tag{6.40}$$

$$\sum_{\nu} b_{\nu}^{3} = \iiint K(\tau_1, \tau_2)\, K(\tau_2, \tau_3)\, K(\tau_3, \tau_1)\, d\tau_1\, d\tau_2\, d\tau_3 \tag{6.41}$$

and so on. Therefore Eq. 6.37 can be expressed in terms of the K's, at least when the b_{ν} are small enough; and there is no problem if the b_{ν} are bigger, for the absolute value of

$$\frac{1}{(1-2ib_{\nu})^{1/2}} \tag{6.42}$$

is

$$\frac{1}{\left(1+4b_{\nu}^{2}\right)^{1/4}} \tag{6.43}$$

and the bigger the b_{ν}, the better the convergence. Hence, expression 6.33 is a parameter that depends on the K, and this parameter is not going to be of serious importance to us in discussing the spectrum. It will merely give a constant factor for the autocorrelation and, therefore, a constant factor for the spectrum that will be the same for all terms. Therefore, we can confine our attention to

$$\begin{aligned} \exp\left(i \iint R(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\right) &= 1 + i \iint R(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha) \\ &\quad - \frac{1}{2!} \iiiint R(\tau_1, \tau_2)\, R(\tau_3, \tau_4)\, dx(\tau_1, \alpha) \ldots dx(\tau_4, \alpha) + \ldots \end{aligned} \tag{6.44}$$

These integrals are homogeneous functionals and can be reduced to the form in which the kernels are symmetrical simply by taking all of the permutations and adding. Having done this, we can build the corresponding G's by the prescription given previously, and we have the development of the expression

$$\exp\left(i \iint K(\tau_1, \tau_2)\, dx(\tau_1, \alpha)\, dx(\tau_2, \alpha)\right) \tag{6.45}$$

in terms of the orthogonal polynomial. You see how I save trouble by using generating functions. I did not have to use them; I could have added up the series directly. But why waste time when there is an easy way? Generating functions are extremely powerful tools for doing a whole bunch of work at the same time.

B. AN ALGEBRA OF CONTINUOUS SYSTEMS

About fifty years ago the mathematician Volterra (1) and others studied a generalization of the linear integral equation. This was

$$x(t) = f(t) + \int_a^t h_1(t, \tau)\, x(\tau)\, d\tau + \int_a^t \int_a^t h_2(t, \tau_1, \tau_2)\, x(\tau_1)\, x(\tau_2)\, d\tau_1\, d\tau_2 + \ldots + \int_a^t \ldots \int_a^t h_n(t, \tau_1, \ldots, \tau_n)\, x(\tau_1) \ldots x(\tau_n)\, d\tau_1 \ldots d\tau_n \qquad (1)$$

In 1942 Wiener (2) applied a similar form to the study of a nonlinear device with a random input. Since then a number of people have studied various aspects of this representation. Brilliant (3), Barrett (4), and Smets (5) have developed various properties and applied the theory to certain nonlinear systems. Cameron and Martin (6), Friedrichs (7), Wiener (8), and Barrett (4) have considered the effects of such systems on random signals.

The work reported here is concerned with systems that are composed of linear subsystems, with memory, which are combined by nonlinear no-memory operations. The operations are basically multiplication operations. We call this class of systems the "Continuous class." It is required that all nonlinear operations shall be expressible in terms of power series.

An algebra is developed for describing these continuous systems, and from this algebra many of the basic properties of the systems can be developed. Representation of the systems, by means of the generalized convolutions previously mentioned, follows from the algebra.

The algebra is a condensed notation which is useful for combining and synthesizing systems. This report is an introduction to the description of Continuous systems by means of this algebra.

[Editor's note: With the author's permission we have substituted underlined Roman letters for the script letters that are customarily used in systems notation.]

1. Basic Definitions and Rules

A system is defined as a physical device that acts upon an input g to produce an output f. Symbolically, it will be represented by an operator $\underline{H}$, which is such that

$$f = \underline{H}[g] \qquad (2)$$

as shown in Fig. VII-5. The input and output are taken as functions of time, so we have

$$f(t) = \underline{H}[g(t)] \qquad (3)$$

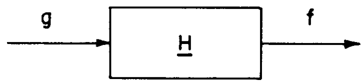

Fig. VII-5.

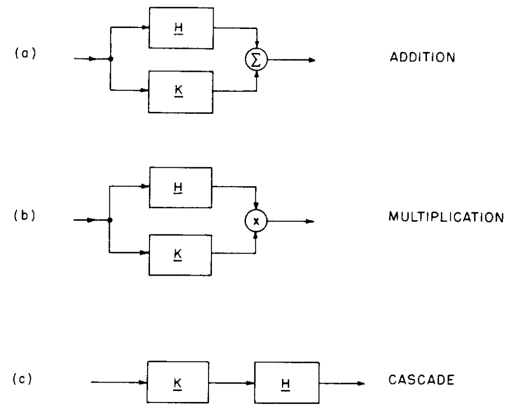

Fig. VII-6.

This algebra is concerned with the combination of such systems. The basic combinations are: (a) addition, denoted by $\underline{H} + \underline{K}$; (b) multiplication, denoted by $\underline{H} \cdot \underline{K}$; (c) cascade, denoted by $\underline{H} * \underline{K}$; as shown in Fig. VII-6. Thus the data for the algebra are: (a) a set of systems, $\underline{H}$, $\underline{K}$, $\underline{L}$, and so forth; (b) a set of operations, +, $\cdot$, and *.

We define ( ) as representing the combination of subsystems into a compound system. From a consideration of the physical nature of combined systems, we have the following basic axioms for the algebra:

A1. $\underline{H} = \underline{H}$ (4)

A2. If $\underline{H} = \underline{A}$ and $\underline{K} = \underline{A}$

then $\underline{H} = \underline{K}$ (5)

For the addition operation:

A3. $\underline{H} + \underline{K} = \underline{K} + \underline{H}$ (6)

A4. $\underline{H} + (\underline{K} + \underline{L}) = (\underline{H} + \underline{K}) + \underline{L}$ (7)

For the multiplication operation:

A5. $\underline{H} \cdot \underline{K} = \underline{K} \cdot \underline{H}$ (8)

A6. $\underline{H} \cdot (\underline{K} \cdot \underline{L}) = (\underline{H} \cdot \underline{K}) \cdot \underline{L}$ (9)

For the cascade operation:

A7. $\underline{H} * (\underline{K} * \underline{L}) = (\underline{H} * \underline{K}) * \underline{L}$ (10)

For combined operations:

C1. $\underline{L} \cdot (\underline{H} + \underline{K}) = \underline{L} \cdot \underline{H} + \underline{L} \cdot \underline{K}$ (11)

C2. $(\underline{H} + \underline{K}) * \underline{L} = \underline{H} * \underline{L} + \underline{K} * \underline{L}$ (12)

C3. $(\underline{H} \cdot \underline{K}) * \underline{L} = (\underline{H} * \underline{L}) \cdot (\underline{K} * \underline{L})$ (13)

It is important to note that, in general,

$$\underline{H} * \underline{K} \neq \underline{K} * \underline{H} \tag{14}$$

$$\underline{L} * (\underline{H} + \underline{K}) \neq \underline{L} * \underline{H} + \underline{L} * \underline{K} \tag{15}$$

$$\underline{L} * (\underline{H} \cdot \underline{K}) \neq (\underline{L} * \underline{H}) \cdot (\underline{L} * \underline{K}) \tag{16}$$

These axioms hold true independently of any particular representation of the operator $\underline{H}$. They are true for all systems.

We shall now specify a class of systems. This class, which we call the Continuous class of systems, will contain all systems that can be represented by

$$\underline{H} = F[\underline{A}_1, \underline{B}_1, \ldots, \underline{M}_1, \ldots] \tag{17}$$

where $\underline{A}_1, \underline{B}_1, \ldots, \underline{M}_1, \ldots$ are linear systems. Here, F represents a function of these linear systems with the operations +, $\cdot$, and $*$. For example,

$$F[\underline{A}_1, \underline{B}_1, \underline{C}_1] = [(\underline{A}_1 + \underline{B}_1) \cdot \underline{C}_1] * \underline{A}_1 \tag{18a}$$

is an explicit representation. A generalization is the class of systems in which $\underline{H}$ appears implicitly:

$$G[\underline{H}, \underline{P}_1, \underline{Q}_1, \ldots, \underline{U}_1, \ldots] = F[\underline{A}_1, \underline{B}_1, \ldots, \underline{M}_1, \ldots] \tag{18b}$$

This class includes systems that involve linear subsystems, together with adders, multipliers, cascade combinations, and nonlinear no-memory elements that are expressible in terms of power series. For example, lumped circuits containing nonhysteretic, nonlinear, but continuous, resistors, capacitors, and inductors are contained in the class. This report deals mainly with the explicit form. One of the most interesting problems is the representation in the explicit form of a system that is defined by the implicit equation.

We shall now show that an ordering can be defined on this (explicit) representation. If we have

$$\underline{H} = \underline{A}_1 \cdot \underline{B}_1 \tag{19}$$

or

$$\underline{H}[x] = \underline{A}_1[x] \cdot \underline{B}_1[x] \tag{20}$$

then

$$\begin{aligned} \underline{H}[\epsilon x] &= \underline{A}_1[\epsilon x] * \underline{B}_1[\epsilon x] \\ &= \epsilon^2 \underline{A}_1[x] \cdot \underline{B}_1[x] \qquad (21) \\ &= \epsilon^2 \underline{H}[x] \end{aligned}$$

since $\underline{A}_1$ and $\underline{B}_1$ are linear systems. Therefore

$$\underline{H} = \underline{A}_1 \cdot \underline{B}_1$$

will be called a pure second-order system, and it will be denoted by

$$\underline{H} = \underline{H}_2 \qquad (22)$$

It is useful to define

$$\underline{H}_2'[xy] = \frac{1}{2} (\underline{A}_1[x] \cdot \underline{B}_1[y] + \underline{A}_1[y] \cdot \underline{B}_1[x]) \qquad (23)$$

Then

$$\underline{H}_2[x] = \underline{H}_2'[xx] = \underline{H}_2'[x^2] \qquad (24)$$

and it is found that $\underline{H}_2'[x^2]$ is linear in x^2. That is,

$$\underline{H}_2'[(x+y)^2] = \underline{H}_2'[x^2] + 2\underline{H}_2'[xy] + \underline{H}_2'[y^2] \qquad (25)$$

Hereafter, the prime will be dropped, and this symmetry can be taken as being implicit in the algebra, so that we have

$$\underline{H}_2[x] = \underline{H}_2'[xx] = \underline{H}_2'[x^2] = \underline{H}_2[x^2] \qquad (26)$$

This equation can be generalized to give

$$\underline{H}_n[x] = \underline{H}_n'[x^n] = \underline{H}_n[x^n] \qquad (27)$$

which is linear in x^n, by taking

$$\underline{H}_n'[x_1 \dots x_n] = \frac{1}{n} \sum \underline{H}_n[x_1 \dots x_n] \qquad (28)$$

where the sum is over all of the ways of arranging $x_1, \dots, x_n$. If $x_1 = x_2 = \dots = x_n$, we shall write

$$\underline{H}_n[x] = \underline{H}_n'[x^n] = \underline{H}_n[x^n] \qquad (29)$$

It can be shown that

$$\underline{H}_m \cdot \underline{H}_n = \underline{H}_{m+n} \qquad (30)$$

and

$$\underline{H}_m * \underline{H}_n = \underline{H}_{mn} \tag{31}$$

which illustrates the effect of elementary operations on this ordering. In general, all systems of this class will have the property

$$\underline{H} = \underline{H}_o + \underline{H}_1 + \underline{H}_2 + \ldots + \underline{H}_n + \ldots = \sum \underline{H}_i \tag{32}$$

The zero-order system, $\underline{H}_o$, is just a constant. In much of the later work it will be assumed that the systems are defined about their dc operating point so that $\underline{H}_o = 0$.

2. Special Systems

Of particular interest is the nonlinear no-memory system. This is denoted algebraically by

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

and described functionally by

$$\begin{aligned} f &= n(x) \\ &= a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots \end{aligned} \tag{34}$$

that is, by a power series. A particular member of this subclass is the identity system $\underline{I}$:

$$\begin{aligned} f &= \underline{I}[x] \\ &= x \end{aligned} \tag{35}$$

The zero system $\underline{O}$ is defined by

$$f = \underline{O}[x] = 0 \tag{36}$$

In algebraic equations, 0 denotes the $\underline{O}$ system

3. Rules for Combining Systems

Let us take

$$\underline{H} = \underline{H}_1 + \underline{H}_2 + \ldots + \underline{H}_n \tag{37}$$

$$\underline{K} = \underline{K}_1 + \underline{K}_2 + \ldots + \underline{K}_m \tag{38}$$

and

$$\underline{L} = \sum \underline{L}_i \tag{39}$$

as the resultant system.

The sum system

$$\underline{L} = \underline{H} + \underline{K} \tag{40}$$

will have

$$\underline{L}_i = \underline{H}_i + \underline{K}_i \tag{41}$$

The product system

$$\underline{L} = \underline{H} \cdot \underline{K} \tag{42}$$

will have

$$\underline{L}_i = \sum_j \underline{H}_j \cdot \underline{K}_{i-j} \tag{43}$$

The cascade system

$$\underline{L} = \underline{H} * \underline{K} \tag{44}$$

gives rise to a complicated expression for $\underline{L}_i$. However,

$$\underline{L} = \sum_j \underline{H}_j * \underline{K} \tag{45}$$

and

$$\underline{H}_j * \underline{K} = \underline{H}_j \left[(\underline{K}_1 + \dots + \underline{K}_m)^j\right] \tag{46}$$

$$= \sum \underline{H}_j [\underline{K}_p \cdot \underline{K}_q \dots \cdot \underline{K}_r] \tag{47}$$

where the summation is the result of expanding

$$(\underline{K}_i + \dots + \underline{K}_m)^j \tag{48}$$

Generally, it seems best to expand $\underline{H}$ in more basic linear subsystems and no-memory nonlinear systems and calculate $\underline{L}$ step by step. Then it will be necessary to compute either

$$\begin{aligned} \underline{L} &= \underline{N} * \underline{K} = \sum \underline{N}_i * \underline{K} \\ &= \sum (\underline{K})^j = \sum (\underline{K}_p \cdot \underline{K}_q \cdot \dots \cdot \underline{K}_r) \end{aligned} \tag{49}$$

as before, and the sum and product rules can be used, or

$$\underline{L} = \underline{A}_1 * \underline{K} \tag{50}$$

in which case

$$\underline{L}_i = \underline{A}_1 * \underline{K}_i \tag{51}$$

4. Generalized Convolutions

A large class of linear systems can be described by the convolution integral. That is,

$$f(t) = \underline{H}_1[x(t)] \tag{52}$$

is equivalent to

$$f(t) = \int h(t-\tau)\, x(\tau)\, d\tau \tag{53}$$

where the integral can be taken from $-\infty$ to $+\infty$. But

$$f(t) = \underline{H}_1[x(t)] \cdot \underline{K}_1[x(t)] \tag{54}$$

is equivalent to

$$f(t) = \left(\int h(t-\tau)\; x(\tau)\, d\tau\right)\left(\int k(t-\tau)\, x(\tau)\, d\tau\right)$$

$$= \iint h(t-\tau_1)\, k(t-\tau_2)\, x(\tau_1)\, x(\tau_2)\, d\tau_1\, d\tau_2 \tag{55}$$

Hence, associated with $\underline{H}_1 \cdot \underline{K}_1$ we have the kernel

$$h(t_1)\; h(t_2) \tag{56}$$

However, in view of the previous symmetrization, and this is where the symmetrization comes in explicitly, the kernel will be taken to be

$$\frac{1}{2}\{h(t_1)\; k(t_2) + h(t_2)\; k(t_1)\} \tag{57}$$

This gives the same result as the unsymmetrical kernel. In general, with a system $\underline{H}_n$ there is associated the kernel

$$h_n(t_1, \ldots, t_n) \tag{58}$$

which is symmetrical in $t_1, \ldots, t_n$. Thus when a system has been described by means of the algebra, the associated kernels can be determined in terms of the kernels of the component linear systems by use of the following relations:

(a) $\underline{H}_n + \underline{K}_n : h(t_1, \ldots, t_n) + k(t_1, \ldots, t_n)$ (59)

(b) $\underline{H}_n \cdot \underline{K}_m : h(t_1, \ldots, t_n)\; k(t_{n+1}, \ldots, t_{n+m})$ (60)

(c) $\underline{H}_1 * (\underline{A}_k \cdot \underline{B}_\ell \cdot \ldots \cdot \underline{C}_m) : \int h_1(\sigma)\, a_k(t_1 - \sigma, \ldots, t_k - \sigma) \ldots c_m(t_{p+1} - \sigma, \ldots, t_{p+m} - \sigma)\, d\sigma$ (61)

where $p + m = k + \ell + \ldots + m$. Relations (b) and (c) are not symmetrical in this form, but they can be symmetrized. The viewpoint taken here is that rather than use the unwieldy general formulas for combining cascade systems, it is better to work through the systems step by step.

5. Canonical Forms

Consider the basic nonlinear operation

$$\underline{H}_1 \cdot \underline{K}_1 \tag{62}$$

where $\underline{H}_1$ and $\underline{K}_1$ are linear. Excluding identity operations $\underline{I}$ and differentiation, a typical term of $h_1(t)$ for lumped systems would be

$$h_1^{(i)}(t) = t^n e^{-a_i t} \tag{63}$$

where a_i can be complex. Then we have

$$\begin{aligned} h_1^{(i)}(t-\sigma) &= (t-\sigma)^n \exp[-a_i(t-\sigma)] \\ &= \left(\sum_{r=0}^{n} \binom{n}{r} t^r \sigma^{n-r} \right) \exp(-a_i t) \exp(+a_i \sigma) \end{aligned} \tag{64}$$

This fact can be used with relation (c) to show that

$$\underline{L}_1 * (\underline{H}_1 \cdot \underline{K}_1) = \sum_k \underline{A}_1^{(k)} \cdot \underline{B}_1^{(k)} \tag{65}$$

where $\underline{L}_1$ is linear, and $\underline{A}_1^{(k)}$ and $\underline{B}_1^{(k)}$ are computed directly by the use of relation (c).

Generalization shows that any system of this class can be represented in canonical form by

$$\underline{H} = \sum \prod \underline{D}_1 \tag{66}$$

that is, by the summation of products of linear systems, and the $\underline{D}_1$ can be computed directly. From this it is possible to show that

$$\underline{H} = \sum \underline{N} * \underline{F}_1 \tag{67}$$

where the $\underline{N}$'s are nonlinear no-memory systems as previously defined, and the $\underline{F}_1$ are linear systems. If the system input is directly into no-memory operations, then

$$\underline{H} = \sum \underline{N}^{(1)} * \underline{F}_1 * \underline{N}^{(2)} \tag{68}$$

Other methods can be used to include $\underline{I}$ operations and differentiation operations, so the canonical form holds for all lumped systems of the Continuous class.

6. Solvability Law

Let us consider the cascade equations

$$\underline{H} * \underline{A} = \underline{B} \tag{69}$$

and

$$\underline{A} * \underline{H} = \underline{B} \tag{70}$$

where $\underline{A}$ and $\underline{B}$ are known systems and $\underline{H}$ is to be determined. This work is directly concerned with certain synthesis problems and also with the inclusion of feedback systems in the Continuous class. This is the basic problem in defining a system that is prescribed by an implicit equation, in the explicit form.

With the algebra developed here it is possible to show that

$$\underline{H} * \underline{A} = \underline{B} \tag{71}$$

defines a unique $\underline{H}$, denoted by

$$\underline{H} = \underline{B} * \underline{A}^{-1} \tag{72}$$

where $\underline{A}^{-1}$ satisfies

$$\underline{A}^{-1} * \underline{A} = \underline{I} \tag{73}$$

$\underline{A}^{-1}$ then also satisfies

$$\underline{A} * \underline{A}^{-1} = \underline{I} \tag{74}$$

Likewise

$$\underline{A} * \underline{H} = \underline{B} \tag{75}$$

is uniquely satisfied by

$$\underline{H} = \underline{A}^{-1} * \underline{B} \tag{76}$$

This leaves three problems unresolved: Does $\underline{H}$ exist? Is $\underline{H}$ stable? Can $\underline{H}$ be described by means of the implicit representation? The results of work on these three questions will be presented in later reports.

D. A. George

References

1. V. Volterra, Leçons sur les Fonctions de Lignes (Gauthier-Villars, Paris, 1913).

2. N. Wiener, Response of a non-linear device to noise, Report 129, Radiation Laboratory, M.I.T., 1942.

3. M. Brilliant, Theory of the analysis of nonlinear systems, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Feb. 1958.

4. J. F. Barrett, The use of functionals in the analysis of non-linear physical systems, Statistical Advisory Unit Report No. 1/57, Ministry of Supply, G. B. (no date given).

5. H. Smets, Etudé des systèmes nonlinèaires par l'analyse fonctionelle, Revue E, Vol. II, No. 3, p. 55 (1957), Sociétè Belge de Électriciens (also, papers on pp. 59 and 63).

6. R. H. Cameron and W. T. Martin, The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals, Ann. Math. 48, 385-389 (1947).

7. F. O. Friedrichs, Integration of functionals, Course Notes, New York University, 1957 (unpublished).

8. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press and John Wiley and Sons, Inc., Cambridge, Mass., and New York, to be published in 1958).

C. THE PREDICTION OF GAUSSIAN-DERIVED SIGNALS

This report deals with the prediction of signals that are derived from signals with Gaussian amplitude distributions by means of certain nonlinear processes. The nonlinear processes belong to the class that can be described by generalized convolutions.

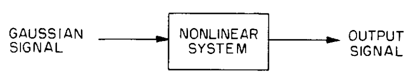

Fig. VII-7.

Fig. VII-8.

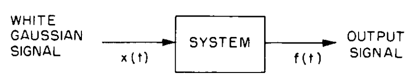

Fig. VII-9.

The process with which we are concerned is represented by Fig. VII-7; we want to predict the output signal. The Gaussian signal is considered as having been obtained from a white Gaussian signal by means of a linear "shaping" filter. The total system is shown in Fig. VII-8. In this report the two cascaded systems will be combined, so that Fig. VII-9 will be the situation that is being considered.

The problem is to find the best estimate of f(t+A), where A is the prediction time, by a realizable operation on f(t).

Symbolically, the system will be represented by an operator $\underline{H}$, where

$$f(t) = \underline{H}[x(t)] \tag{1}$$

The procedure used will be to determine a realizable operation $\underline{H}_A$ so that

$$\tilde{f}(t+A) = \underline{H}_A[x(t)] \tag{2}$$

where $\tilde{f}(t+A)$ is the best estimate of $f(t+A)$ in the mean-square sense. Then the inverse system $\underline{H}^{-1}$, which is such that

$$\underline{H}^{-1}[\underline{H}[x(t)]] = x(t) \tag{3}$$

will be found. It is seen that the cascade combination of $\underline{H}_A$ and $\underline{H}^{-1}$ gives the optimum predictor $\underline{P}$. That is,

$$\tilde{f}(t+A) = \underline{P}[f(t)] = \underline{H}_A[\underline{H}^{-1}[f(t)]] \tag{4}$$

To represent the operation of the system $\underline{H}$, generalized convolutions will be used. Then

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

$$+ \dots + \int_{-\infty}^{t} \dots \int_{-\infty}^{t} h_m(t-\tau_1, \dots, t-\tau_m)\, x(\tau_1) \dots x(\tau_m)\, d\tau_1 \dots d\tau_m$$

$$= \underline{H}[x(t)] \tag{5}$$

The following notation will be used:

$$f(t) = \underline{H}[x(t)] = \sum_{n=1}^{m} \underline{H}_n[x(t)] = \sum_{n=1}^{m} f_n(t) \tag{6}$$

where

$$f_n(t) = \int_{-\infty}^{t} \dots \int H_n(t-\tau_1, \dots, t-\tau_n)\, x(\tau_1) \dots x(\tau_n)\, d\tau_1 \dots d\tau_n = \underline{H}_n[x(t)] \tag{7}$$

The kernels $\underline{H}_n(t_1, \dots, t_n)$ may always be taken as symmetrical in $t_1, \dots, t_n$.

1. Prediction

Now we shall deal with the determination of $\underline{H}_A$. The output of the n^{th} order term at time t is

$$f_n(t) = \int_{-\infty}^{t} \dots \int h_n(t-\tau_1, \dots, t-\tau_n)\, x(\tau_1) \dots x(\tau_n)\, d\tau_1 \dots d\tau_n \tag{8}$$

and at time t + A it is

$$f_n(t+A) = \int_{-\infty}^{t} \cdots \int h_n(t+A-\tau_1, \ldots, t+A-\tau_n)\, x(\tau_1) \ldots x(\tau_n)\, d\tau_1 \ldots d\tau_n$$

$$= \sum_{r=0}^{n} \binom{n}{r} \underbrace{\int_{-\infty}^{t} \cdots \int}_{\substack{n-r\\ \text{integrals}}} \underbrace{\int_{t}^{t+A} \cdots \int}_{r\ \text{integrals}} h_n(t+A-\tau_1, \ldots, t+A-\tau_n) x(\tau_1) \ldots x(\tau_n) d\tau_1 \ldots d\tau_n \tag{9}$$

by expansion and use of symmetry. Now consider a single term

$$f_{n,r}(t+A) = \underbrace{\int_{-\infty}^{t} \cdots \int}_{n-r} \underbrace{\int_{t}^{t+A} \cdots \int}_{r} h_n(t+A-\tau_1, \ldots, t+A-\tau_n)\, x(\tau_1) \ldots x(\tau_n)\, d\tau_1 \ldots d\tau_n \tag{10}$$

the r variables $x(\tau_1) \ldots x(\tau_r)$ integrated over the time range $[t, t+A]$ are integrated over the future. Therefore, to find the best mean-square estimate, $\tilde{f}_{n,r}(t+A)$, of $f_{n,r}(t+A)$, these variables are averaged. That is,

$$\tilde{f}_{n,r}(t+A) = \underbrace{\int_{-\infty}^{t} \cdots \int}_{n-r} \underbrace{\int_{t}^{t+A} \cdots \int}_{r} h_n(t+A-\tau_1, \ldots, t+A-\tau_n)\, \left[\overline{x(\tau_1) \ldots x(\tau_r)}\right]$$

$$x(\tau_{r+1}) \ldots x(\tau_n)\, d\tau_1 \ldots d\tau_n \tag{11}$$

where $\left[\overline{x(\tau_1) \ldots x(\tau_r)}\right]$ is the average of $[x(\tau_1) \ldots x(\tau_r)]$. It follows from Wiener (1) that — (at least when the integrals exist),

$$\tilde{f}_{n,r}(t+A) = 0 \qquad \text{for } r \text{ odd}$$

$$= (r-1)(r-3) \ldots 1 \underbrace{\int_{-\infty}^{t} \cdots \int_{-\infty}^{t}}_{n-r} \left[\underbrace{\int_{0}^{A} \cdots \int_{0}^{A}}_{r} h_n(\sigma_1, \sigma_1, \ldots \sigma_{r/2}, \sigma_{r/2}, t-\tau_1+A \right.$$

$$\left. \ldots, t-\tau_{n-r}+A) \right] x(\tau_1) \ldots x(\tau_{n-r})\, d\tau_1 \ldots d\tau_{n-r} \qquad \text{for } r \text{ even} \tag{12}$$

That is, a new kernel $h^{(r)}_{A,n}$ has been developed, such that

$$\begin{aligned} h^{(r)}_{A,n}(\tau_1, \dots, \tau_{n-r}) &= 0 \quad \text{for } r \text{ odd and all } \tau_i, \text{ where } i = 1, \dots, n-r \\ &= (r-1)(r-3)\dots 1 \int_0^A \int_0^A h_n(\sigma_1, \sigma_1, \dots \sigma_{r/2}, \sigma_{r/2}, \tau_1 + A, \\ &\qquad \dots, \tau_{n-r} + A)\, d\sigma_1 \dots d\sigma_{r/2} \quad \text{for } r \text{ even and all } \tau_i \geq 0 \\ &= 0 \quad \text{for } r \text{ even and some } \tau_i < 0 \end{aligned} \tag{13}$$

This kernel describes a system $\underline{H}^{(r)}_{A,n}$ which is such that

$$\tilde{f}_{n,r}(t+A) = \underline{H}^{(r)}_{A,n}[x(t)] \tag{14}$$

and so

$$\tilde{f}_n(t+A) = \sum_{r=0}^{n} \binom{n}{r} \underline{H}^{(r)}_{A,n}[x(t)] \tag{15}$$

Since

$$\tilde{f}(t+A) = \sum_{n=1}^{m} \tilde{f}_n(t+A) \tag{16}$$

then

$$\tilde{f}(t+A) = \underline{H}_A[x] \tag{17}$$

where

$$\underline{H}_A[x(t)] = \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{n}{r} \underline{H}^{(r)}_{A,n}[x(t)] \tag{18}$$

Thus, $\underline{H}_A$ has been prescribed.

2. Synthesis

The synthesis of the predictor is divided into two parts: the system $\underline{H}_A$ and the inverse system $\underline{H}^{-1}$. We shall first consider $\underline{H}^{-1}$.

Formally, at least, the circuit of Fig. VII-10 can be used to obtain the inverse system. But since there are some unresolved problems of stability connected with this system, the following discussion will be limited to a special case. In this case

$$\underline{H} = \underline{N} * \underline{L} \tag{19}$$

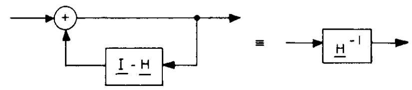

Fig. VII-10.

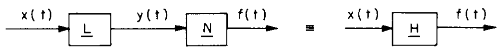

Fig. VII-11.

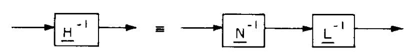

Fig. VII-12.

where $\underline{L}$ is a linear system, $\underline{N}$ is a no-memory nonlinear system, and * denotes cascade. This is shown in Fig. VII-11. The system $\underline{N}$ is specified by

$$
\begin{aligned}
f(t) &= \underline{N}[y(t)] = n(y(t)) \\
&= a_1 y(t) + a_2 y^2(t) + \dots + a_m y^m(t)
\end{aligned} \tag{20}
$$

If the function n specifies a one-to-one relationship between x and y, there exists an inverse function n^{-1} and a corresponding inverse system $\underline{N}^{-1}$. Then

$$
\underline{H}^{-1} = (\underline{N} * \underline{L})^{-1} = \underline{L}^{-1} * \underline{N}^{-1} \tag{21}
$$

as shown in Fig. VII-12. If $\underline{L}^{-1}$, the inverse of the linear system $\underline{L}$, is stable, then $\underline{H}^{-1}$ is stable. $\underline{L}^{-1}$ may be found by the usual linear techniques, and $\underline{N}^{-1}$ is a well-defined no-memory operation.

Now let us consider the system $\underline{H}_A$. The n^{th} order system $\underline{H}_n$ has a kernel

$$
h_n(\tau_1, \dots, \tau_n) = a_n\, \ell(\tau_1) \dots \ell(\tau_n) \tag{22}
$$

where $\ell(t)$ is the impulse response of the linear system $\underline{L}$. The corresponding kernel $h_{A,n}^{(r)}$ is given by

$$
\begin{aligned}
h_{A,n}^{(r)}(\tau_1, \dots, \tau_{n-r}) = (r-1)(r-3)\dots 1 &\left\{ a_n \int_0^A \dots \int_0^A \ell^2(\sigma_1)\dots \ell^2(\sigma_{r/2}) d\sigma_1 \dots d\sigma_{r/2} \right\} \\
&\cdot \ell(\tau_1 + A) \dots \ell(\tau_{n-r} + A) \quad \text{for all } \tau_i \geq 0 \text{ and } r \text{ even}
\end{aligned} \tag{23}
$$

and by

$$
h_{A,n}^{(r)}(\tau_1, \dots, \tau_{n-r}) = 0 \quad \text{for } r \text{ odd or some } \tau_i < 0 \tag{24}
$$

So $\ell(\tau_1 + A)$ can be synthesized as in the linear predictor and $\underline{H}_{A,n}^{(r)}$ synthesized by

$$\underline{H}_{A,n}^{(r)} = \underline{M}_{n-r} * \underline{L}_A \tag{25}$$

where $\underline{M}_{n-r}$ is a nonlinear no-memory system specified by

$$\underline{M}_{n-r}[y] = k_{n,r} y^{n-r} \tag{26}$$

where

$$k_{n,r} = (r-1)(r-3)\dots 1 \left\{ a_n \int_0^A \dots \int_0^A \ell^2(\sigma_1)\dots\ell^2(\sigma_{r/2}) d\sigma_1 \dots d\sigma_{r/2} \right\} \text{ for } r \text{ even}$$

$$= 0 \text{ for } r \text{ odd} \tag{27}$$

and $\underline{L}_A$ is the linear system corresponding to $\ell(\tau+A)$, which is zero for $\tau < 0$.

Since

$$\underline{H}_A[x(t)] = \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{n}{r} H_{A,n}^{(r)}[x(t)] \tag{28}$$

the system $\underline{H}_A$ becomes

$$\underline{H}_A = \underline{M} * \underline{L}_A \tag{29}$$

where $\underline{M}$ is the nonlinear no-memory system resulting from the summation of all the $\underline{M}_{n-r}$ systems by the above equation. That is,

$$\underline{M} = \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{n}{r} \underline{M}_{n-r} \tag{30}$$

Therefore the predictor $\underline{P}$ has been synthesized, and

$$\underline{P} = \underline{H}_A * \underline{H}^{-1}$$

$$= \underline{M} * \underline{L}_A * \underline{L}^{-1} * \underline{N}^{-1} \tag{31}$$

This approach can be extended to systems of the form

$$\underline{H} = \underline{N}^{(1)} * \underline{L}^{(1)} * \underline{N}^{(2)} * \underline{L}^{(2)} * \dots * \underline{N}^{(k)} * \underline{L}^{(k)} \tag{32}$$

where the $\underline{N}^{(i)}$ are invertible nonlinear no-memory systems, and the $\underline{L}^{(i)}$ are linear systems with memory. In some cases, the systems

$$\underline{H} = \underline{H}^{(1)} + \underline{H}^{(2)} + \dots + \underline{H}^{(h)} \tag{33}$$

can be handled. The $\underline{H}^{(i)}$ are of the cascade form shown in Eq. 32.

3. Example

Let us consider

$$\underline{H} = \underline{N} * \underline{L} \tag{34}$$

where

$$\underline{N}[y] = n(y) = y + y^3 = f \tag{35}$$

and

$$\ell(t) = \frac{\exp(-\alpha t) - \exp(-\beta t)}{\beta - \alpha} \quad \text{for } t \geq 0 \tag{36}$$

Then the inverse function $n^{-1}(f)$ is well defined, and $\underline{N}^{-1}$ is specified. The transform of $\ell(t)$ is

$$L(s) = \frac{1}{(s+\alpha)(s+\beta)} \tag{37}$$

and the inverse linear system, $\underline{L}^{-1}$ with impulse response h(t), has the transform

$$H(s) = (s+\alpha)(s+\beta) \tag{38}$$

Associated with $\underline{L}_A$ is the transform

$$L_A(s) = e^{sA} \int_A^\infty \left(\frac{\exp(-\alpha t) - \exp(-\beta t)}{\beta - \alpha}\right) \exp(-st)\, dt$$

$$= \frac{(\beta \exp(-\alpha A) + \alpha \exp(-\beta A)) + (\exp(-\alpha A) - \exp(-\beta A))\, s}{(\beta-\alpha)(s+\alpha)(s+\beta)} \tag{39}$$

All the $k_{n,r}$ are zero except

$$k_{3,2} = \int_0^A \ell^2(\sigma)\, d\sigma = \frac{1}{(\beta-\alpha)^2} \left\{\frac{1 - \exp(-2\alpha A)}{2\alpha} + \frac{1 - \exp(-2\beta A)}{2\beta} - \frac{1 - \exp[-(\alpha+\beta)A]}{\alpha + \beta}\right\} \tag{40}$$

$$k_{1,o} = k_{3,o} = 1 \tag{41}$$

Therefore

$$\underline{M}_{3-2}[y] = k_{3,2}\, y \tag{42}$$

$$\underline{M}_{1-o}[y] = y \tag{43}$$

$$\underline{M}_{3-o}[y] = y \tag{44}$$

and

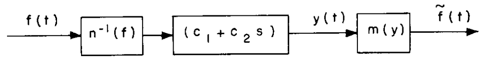

Fig. VII-13.

$$\underline{M}[y] = (1 + 3k_{3,2})y + y^3 = m(y) \tag{45}$$

And hence $\underline{M}$ is specified. The system $\underline{L}_A * \underline{L}^{-1}$ has the transform

$$L_A(s) \cdot H(s) = \frac{(\beta \exp(-\alpha A) + \alpha \exp(-\beta A)) + (\exp(-\alpha A) - \exp(-\beta A))\, s}{(\beta-\alpha)}$$

$$= c_1 + c_2 s \tag{46}$$

and the predictor $\underline{P}$ is shown in Fig. VII-13.

D. A. George

References

1. N. Wiener, Nonlinear Problems in Random Theory (The Technology Press and John Wiley and Sons, Inc., Cambridge, Mass., and New York, to be published in 1958).

D. A SAMPLING THEOREM FOR STATIONARY RANDOM PROCESSES*

Let $[x(t)]$ be a stationary process and $[f(t)]$ a process generated by samples of $[x(t)]$. We wish to answer the following question:

Given samples of $x(t)$, taken uniformly at a rate $1/T_o$ over the time interval $(-NT_o, NT_o)$, what interpolatory function, $s(t)$, gives us best agreement between $f(t)$ – the interpolated sample function – and $x(t)$ during the time interval $(-NT_o, NT_o)$ for all members of the ensemble $[x(t)]$?

By "best agreement during the time interval $(-NT_o, NT_o)$ for all members of the ensemble $[x(t)]$" we mean that we want to minimize

$$I = \frac{1}{2NT_o} \int_{-NT_o}^{NT_o} E \cdot \left[[x(t) - f(t)]^2 \right]^{**} dt$$

where

*This work began in discussions with W. R. Bennett, M. Karnaugh, and H. P. Kramer, of Bell Telephone Laboratories, where part of the following analysis was carried out by the writer.

**E[ ] = ensemble average of [ ].

$$f(t) = \sum_{n=-N}^{N} x(nT_o)\, s(t - nT_o)$$

We shall now obtain a more explicit expression of our criterion.

$$[x(t) - f(t)]^2 = x^2(t) - 2 \sum_{m=-N}^{N} x(t) \cdot x(mT_o) \cdot s(t - mT_o)$$

$$+ \sum_{m,n=-N}^{N} x(mT_o) \cdot x(nT_o) \cdot s(t - mT_o) \cdot s(t - nT_o)$$

If $\phi_x(\tau)$ is the autocorrelation of $[x(t)]$, we have

$$E \cdot \left[[x(t) - f(t)]^2 \right] = E[x^2(t)] - 2 \sum_{m=-N}^{N} \phi_x(t - mT_o) \cdot s(t - mT_o)$$

$$+ \sum_{m,n=-N}^{N} \phi_x([m-n]T_o) \cdot s(t - mT_o) \cdot s(t - nT_o)$$

Since $[x(t)]$ is assumed stationary, $E[x^2(t)]$ is a constant, which we shall call C. The expression to be minimized becomes

$$I = \frac{1}{2NT_o} \int_{-NT_o}^{NT_o} \left[C - 2 \sum_{m=-N}^{N} \phi_x(t - mT_o)\, s(t - mT_o) \right.$$

$$\left. + \sum_{m=-N}^{N} \sum_{n=+N}^{-N} \phi_x([m-n]T_o) \cdot s(t - mT_o) \cdot s(t - nT_o) \right] dt \tag{1}$$

Next we change summation index, letting $m - n = k$.

$$I = \frac{1}{2NT_o} \int_{-NT_o}^{NT_o} \left[C - 2 \sum_{m=-N}^{N} \phi_x(t - mT_o)\, s(t - mT_o) \right.$$

$$\left. + \sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} \phi_x(kT_o)\, s(t - mT_o)\, s(t - mT_o + kT_o) \right] dt \tag{2}$$

The order of summation can be interchanged by the formula

$$\sum_{m=-N}^{N} \sum_{k=m-N}^{m+N} = \sum_{k=-2N}^{0} \sum_{m=-N}^{k+N} + \sum_{k=1}^{2N} \sum_{m=k-N}^{N}$$

We also allow the limits of integration to become infinite by introducing a function, $g_N(t)$, in the integrand; $g_N(t)$ is 1 in the time interval $(-NT_o, NT_o)$ and is zero elsewhere.

Interchanging orders of summation and integration, we have

$$I = \frac{1}{2NT_o} \sum_{m=-N}^{N} \int_{-\infty}^{\infty} g_N(t) \left[C - 2\phi_x(t - mT_o)\, s(t - mT_o)\right] dt$$

$$+ \sum_{k=-2N}^{0} \sum_{m=-N}^{k+N} \int_{-\infty}^{\infty} \phi_x(kT_o) \cdot s(t - mT_o) \cdot s(t - mT_o + kT_o)\, g_N(t)\, dt$$

$$+ \sum_{k=1}^{2N} \sum_{m=k-N}^{N} \int_{-\infty}^{\infty} \phi_x(kT_o)\, s(t - mT_o)\, s(t - mT_o + kT_o)\, g_N(t)\, dt \tag{3}$$

We now make a change of variable, $u = t - mT_o$, and interchange the orders of summation and integration.

$$I = \frac{1}{2NT_o} \int_{-\infty}^{\infty} \left\{ \left[C - 2\phi_x(u) \cdot s(u)\right] \cdot \sum_{m=-N}^{N} g_N(u + mT_o) \right.$$

$$+ \left[\sum_{k=-2N}^{0} \phi_x(kT_o)\, s(u + kT_o)\right] \cdot s(u) \cdot \sum_{m=-N}^{k+N} g_N(u + mT_o)$$

$$\left. + \left[\sum_{k=1}^{2N} \phi_x(kT_o)\, s(u + kT_o)\right] \cdot s(u) \cdot \sum_{m=k-N}^{N} g_N(u + mT_o) \right\} du \tag{4}$$

The expression

$$f(u) = \sum_{m=-N}^{N} g_N(u + mT_o)$$

can be described as follows.

$$
\begin{aligned}
f(u) &= 2N \text{ for } |u| \leqslant T_o \\
&= (2N-1) \text{ for } T_o < |u| \leqslant 2\,T_o \\
&\quad\vdots \\
&= 1 \text{ for } (2N-1)\,T_o < |u| \leqslant 2\,NT_o \\
&= 0 \text{ elsewhere}
\end{aligned}
$$

Let

$$
\begin{aligned}
f_k(u) &= \sum_{m=-N}^{k+N} g_N(u+mT_o) \text{ for } -2N \leqslant k \leqslant 0 \\
&= \sum_{m=k-N}^{N} g_N(u+mT_o) \text{ for } 1 \leqslant k \leqslant 2N \\
&= 0 \text{ for all other } k
\end{aligned}
$$

Note that $f_o(u) = f(u)$, and $f_k(u-kT_o) = f_{-k}(u)$.

We can rewrite Eq. 4 as

$$
I = \frac{1}{T_o}\int_{-\infty}^{\infty}\left\{\left[C - 2\phi_x(u)\;s(u)\right]\cdot\frac{1}{2N}f(u) + s(u)\sum_{k=-2N}^{2N}\phi_x(kT_o)\cdot s(u+kT_o)\cdot\frac{1}{2N}\cdot f_k(u)\right\}du \tag{5}
$$

A necessary condition for an unconstrained minimum for I (by varying s) is

$$
-2\phi_x(u)\cdot\frac{f(u)}{2N} + \sum_{k=-2N}^{2N}\phi_x(kT_o)\cdot s(u+kT_o)\cdot\left[\frac{f_k(u)}{2N} + \frac{f_k(u-kT_o)}{2N}\right] = 0 \tag{6}
$$

Recalling that $f_k(u-kT_o) = f_{-k}(u)$, we can rewrite Eq. 6 as follows:

$$
-2\phi_x(u)\cdot\frac{f(u)}{2N} + \sum_{k=-2N}^{2N}\phi_x(kT_o)\cdot s(u+kT_o)\left[\frac{f_k(u)+f_{-k}(u)}{2N}\right] = 0 \tag{7}
$$

Since $f(u)$, $f_k(u)$, and $f_{-k}(u)$ are all zero for $|u| > 2NT_o$, $s(u)$ is arbitrary for

$|u| > 2NT_o$. For simplicity we shall let s(u) be zero for $|u| > 2NT_o$.

Let us now rewrite Eq. 7 for the special case in which $\phi_x(kT_o) = 0$ for $k = \pm 1, \pm 2, \pm 3, \ldots, \pm 2N$. We then have

$$-2\phi_x(u) \cdot \frac{f(u)}{2N} + \phi_x(0) \cdot s(u) \left[\frac{2f(u)}{2N}\right] = 0$$

A solution of this equation is

$$s(u) = 0 \text{ for } |u| > 2NT_o$$

$$s(u) = \frac{\phi_x(u)}{\phi_x(0)} \text{ for } |u| \leq 2NT_o$$

It can be shown that as N approaches infinity, $\frac{f(u)}{2N}$ approaches 1, and $\frac{f_k(u)}{2N}$ approaches 1. More explicitly, given values of u, k, and $\epsilon > 0$, there exists an N_1, which is such that

$$1 - \frac{f(u)}{2N} < \epsilon \text{ and } 1 - \frac{f_k(u)}{2N} < \epsilon$$

for all $N > N_1$.

Thus, in the limiting case in which our sampling and reconstructing intervals become infinite, Eq. 7 becomes

$$-2\phi_x(u) + 2 \sum_{k=-\infty}^{\infty} \phi_x(kT_o) \cdot s(u + kT_o) = 0$$

If $\phi_x(u)$ is band-limited to the frequency interval (-W, W) and the sequence $\phi_x(kT_o)$ represents samples of $\phi_x(u)$ taken at the Nyquist rate, Shannon's sampling theorem (1) tells us that $s(u) = (\sin 2\pi Wu)/(2\pi Wu)$.

To illustrate our theory, let us consider the case in which

$$\phi_x(\tau) = 2 - \frac{\tau}{T_o} \text{ for } 0 \leq |\tau| \leq 2T_o$$

$$= 0 \text{ elsewhere}$$

Equation 7 can then be rewritten as follows,* letting N=1,

$$4 - 2\frac{u}{T_o} = \frac{3}{2} s(T_o - u) + 4s(u) + \frac{3}{2} s(u + T_o) \tag{8a}$$

for $0 \leq u \leq T_o$

*The form of Eq. 7 indicates that we should look for a solution that is symmetric in u; i.e., we replace $s(u - T_o)$ by $s(T_o - u)$.

$$2 - \frac{u_1}{T_o} = \frac{1}{2} s(u_1 - T_o) + 2s(u_1) \tag{8b}$$

for $T_o < u_1 \leq 2T_o$

We can rewrite Eq. 8b in the form

$$1 - \frac{u}{T_o} = \frac{1}{2} s(u) + 2s(u + T_o) \tag{9}$$

for $0 < u \leq T_o$

Solving Eqs. 8a and 9 simultaneously yields

$$\left.\begin{aligned} s(u) &= 0.804 - 0.588 \frac{u}{T_o} \\ s(u + T_o) &= 0.299 - 0.353 \frac{u}{T_o} \end{aligned}\right\} \quad 0 < u < T_o \tag{10}$$

Substituting the values $u = 0$ and $u = T_o$ in Eq. 8a and $u = T_o$ in Eq. 9 gives us the system:

$$\begin{aligned} 4 &= 4s(0) + 3s(T_o) \\ 2 &= \frac{3}{2} s(0) + 4s(T_o) + \frac{3}{2} s(2T_o) \\ 0 &= \frac{1}{2} s(T_o) + 2s(2T_o) \end{aligned} \tag{11}$$

The solution of this system, $[s(0) = 0.85,\ s(T_o) = 0.20,\ s(2T_o) = -0.05]$, does not agree with the corresponding extrapolated values of Eqs. 10.

Since we would probably want our approximation to $x(t)$ to be as smooth as possible, the extrapolated solutions of Eqs. 10 should be used. As we remarked after Eq. 7, $s(u) = 0$ for $|u| > 2T_o$.

Our interpolatory function $s(u)$ may be constrained to be square-integrable and to be zero for $u < 0$ by straightforward modification of Eq. 7.

D. W. Tufts

References

1. C. E. Shannon, Proc. IRE 37, 10-21 (Jan. 1949).

E. AN ERGODIC THEOREM FOR A CLASS OF RANDOM PROCESSES

We shall consider an ensemble of real signals $[m(t)]$ with the following properties:

a. The sequence of functions

$$\{g_N(t)\} = \left\{\frac{1}{2N+1}\sum_{n=-N}^{N} m_p(t+nT_o)\cdot m_p(t+nT_o+\tau)\right\}$$

is uniformly bounded and

$$\lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} m_p(t+nT_o)m_p(t+nT_o+\tau) = E[m(t)\cdot m(t+\tau)]$$

for all t and τ and for almost all $m_p(t)$. $E[m(t)\cdot m(t+\tau)]$ is the ensemble average of $m(t)\cdot m(t+\tau)$; $m_p(t)$ is a particular member of the ensemble $[m(t)]$; and T_o and A are real constants, characteristic of the process. N may take on any of the values 0, 1, 2, . . .

b. $$\lim_{\substack{T\to\infty\\ A\to\infty}}\frac{1}{A+T}\int_{-A}^{T} m_A(t)\cdot m_A(t+\tau)\,dt = \phi_m(\tau)$$

exists for all τ and almost all $m_A(t)$ and is independent of the sequence of values through which $T\to\infty$ and $A\to\infty$. Such an ensemble will be called cyclo-ergodic. (This name was suggested by W. R. Bennett, of Bell Telephone Laboratories, with whom the writer discussed the following result.)

Theorem: If $[m(t)]$ is cyclo-ergodic, the ensemble average of $m(t)\cdot m(t+\tau)$ is related to the time average of $m_p(t)\cdot m_p(t+\tau)$ as follows: For all real C and all positive integers K,

$$\frac{1}{KT_o}\int_{C}^{C+KT_o} E[m(t)\cdot m(t+\tau)]\,dt = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} m_p(t)m_p(t+\tau)\,dt$$

Proof: Since $[m(t)]$ is cyclo-ergodic,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} m_p(t)m_p(t+\tau)\,dt = \lim_{N\to\infty}\frac{1}{(2N+1)KT_o}\int_{-NKT_o+C}^{(N+1)KT_o+C} m_p(t)m_p(t+\tau)\,dt$$

where N is a positive integer.

Next we replace the integral by a sum of integrals to obtain

$$\lim_{N\to\infty}\frac{1}{(2N+1)KT_o}\sum_{n=-N}^{N}\int_{nKT_o+C}^{(n+1)KT_o+C} m_p(t)m_p(t+\tau)\,dt$$

We now make the change of variable $u = t - nKT_o$

$$\lim_{N\to\infty} \frac{1}{(2N+1)KT_o} \sum_{n=-N}^{N} \int_C^{C+KT_o} m_p(u+nKT_o)\, m_p(u+nKT_o+\tau)\, du$$

$$= \lim_{N\to\infty} \frac{1}{KT_o} \int_C^{C+KT_o} \frac{1}{(2N+1)} \sum_{n=-N}^{N} m_p(t+nKT_o)\, m_p(t+nKT_o+\tau)\, dt$$

A theorem of Lebesgue states that if the sequence $g_N(t)$ is uniformly bounded in S, and if $\lim_{N\to\infty} g_N(t) = g(t)$ exists almost everywhere in S, we have

$$\lim_{N\to\infty} \int_S g_N(t)\, dt = \int_S g(t)\, dt$$

In our case, we have

$$g_N(t) = \frac{1}{2N+1} \sum_{n=-N}^{N} m_p(t+nKT_o)\, m_p(t+nKT_o+\tau)$$

and $g_N(t)$ is uniformly bounded on the interval $C \leq t \leq C + KT_o$ because, by property a, $|g_N(t)| \leq A$ for all N and for all t and τ.

Also from property a we have

$$\lim_{N\to\infty} g_N(t) = E[m(t) \cdot m(t+\tau)]$$

Thus we can interchange the orders of $\lim_{N\to\infty}$ and integration to obtain

$$\frac{1}{KT_o} \int_C^{C+KT_o} E[m(t) \cdot m(t+\tau)]\, dt$$

D. W. Tufts

F. OPTIMUM NONLINEAR FILTERS WITH FIXED OUTPUT-NETWORKS

A class of nonlinear filters was described by A. G. Bose in Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956. The purpose of the present work is to extend the optimization procedure for these filters to some cases in which the output of a fixed network is to be optimized by the selection of a filter that precedes the network, as in Fig. VII-14. This problem was suggested by Professor Bose. The three types of fixed networks that we have considered are: linear networks with memory, nonlinear, no-memory networks, and networks with inverses.

The general form of the filter described by Bose is shown in Fig. VII-15. The C's

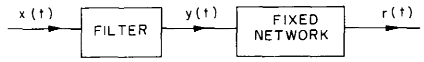

Fig. VII-14. Filter with fixed output-network.

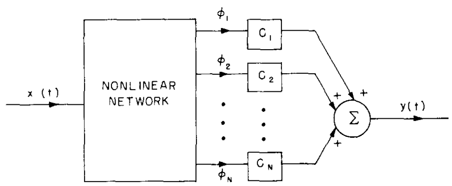

Fig. VII-15. Nonlinear filter.

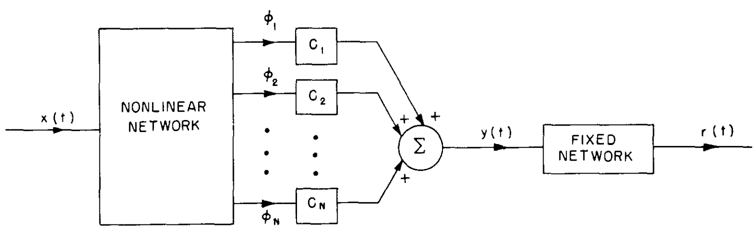

Fig. VII-16. Nonlinear filter of Fig. VII-15 with fixed output-network.

are gains. If y(t) is the filter output and z(t) the desired output, then the C's are set so that the mean-square error

$$\overline{E^2(t)} = \overline{[y(t) - z(t)]^2} \tag{1}$$

is minimized. The bars indicate time averages. These optimum C's are to be determined by measurements in a manner whereby each C can be determined independently of all other C's. Bose showed that a sufficient condition for this optimization procedure is that the ϕ's, which are the outputs of the nonlinear network in Fig. VII-15, satisfy the equation

$$\overline{\phi_i(t)\, \phi_j(t)} = 0 \qquad i \neq j \tag{2}$$

We next consider the question of sufficient conditions for the determination, by independent measurements, of optimum C's for the case in which the filter is cascaded with a fixed output-network. The complete system is given by Fig. VII-16. The mean-square error, which is to be minimized, is

$$\overline{E^2(t)} = \overline{[r(t) - z(t)]^2} \tag{3}$$

where r(t) is the system output. Let us define the functional F describing the fixed network by

$$r(t) = F\left(\sum_{n=1}^{N} C_n \phi_n\right) \tag{4}$$

There are two conditions which, together, are sufficient for determining optimum C's by independent measurements. They are (a) that $r(t)$ can be written as

$$r(t) = \sum_{n=1}^{N} F(C_n \phi_n) \tag{5}$$

and (b) that

$$\overline{F(C_i \phi_i)\, F(C_j \phi_j)} = 0 \qquad i \neq j \tag{6}$$

The sufficiency of these two conditions can be shown as follows. Substitution of Eq. 5 in Eq. 3 gives

$$\overline{E^2(t)} = \overline{\left[\sum_{n=1}^{N} F(C_n \phi_n) - z(t)\right]^2} \tag{7}$$

Expanding the right-hand side, we get

$$\overline{E^2(t)} = \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{F(C_i \phi_i)\, F(C_j \phi_j)} - 2 \sum_{n=1}^{N} \overline{F(C_n \phi_n)\, z(t)} + z^2(t) \tag{8}$$

The orthogonality condition given by Eq. 6 makes the cross terms in the double sum zero. Hence we have

$$\overline{E^2(t)} = \sum_{n=1}^{N} \overline{F^2(C_n \phi_n)} - 2 \sum_{n=1}^{N} \overline{F(C_n \phi_n)\, z(t)} + \overline{z^2(t)} \tag{9}$$

This equation can be rewritten as

$$\overline{E^2(t)} = \sum_{n=1}^{N} \overline{[F(C_n \phi_n) - z(t)]^2} - (N-1)\, \overline{z^2(t)} \tag{10}$$

It is important to note that Eq. 10 gives the mean-square error as a sum of mean-square errors, each of which is a function of only one C.

From this analysis it is clear that the procedure for determining the C's is to measure the mean-square error as a function of one of the C's while holding all the other C's constant. The optimum value for any C is the value that minimizes the mean-square error when that C is being varied.

1. Linear Fixed Networks with Memory

If the fixed network is linear, with impulse response h(t), the sufficient conditions given by Eqs. 5 and 6 can be satisfied. Equation 5 is satisfied because of the linearity of the fixed network. Equation 6 can be satisfied by forming a set ϕ^* that is composed of linear combinations of ϕ's; that is,

$$\phi_n^*(t) = b_n \left[\phi_n(t) + \sum_{j=1}^{n-1} a_{jn}\, \phi_j^*(t)\right] \tag{11}$$

where the ϕ^*'s satisfy the equation

$$\int h(\tau_1)\, \phi_j^*(t-\tau_1)\, d\tau_1 \int h(\tau_2)\, \phi_k^*(t-\tau_2)\, d\tau_2 = \left.\begin{matrix} 1 & j=k \\ 0 & j\neq k \end{matrix}\right\} \tag{12}$$

To solve for any constant a_{kn} of the a_{jn}'s in Eq. 11, we apply the orthogonality condition of Eq. 12 to Eq. 11. In so doing we have

$$\overline{\int h(\tau_1)\, \phi_k^*(t-\tau_1)\, d\tau_1 \int h(\tau_2) b_n \left[\phi_n(t-\tau_2) + \sum_{j=1}^{n-1} a_{jn}\, \phi_j^*(t-\tau_2)\right] d\tau_2} = 0 \tag{13}$$

Applying Eq. 12 to the terms in the sum, we get

$$a_{kn} = - \overline{\int h(\tau_1)\, \phi_k^*(t-\tau_1)\, d\tau_1 \int h(\tau_2)\, \phi_n(t-\tau_2)\, d\tau_2} \tag{14}$$

The term on the right-hand side of this equation can be determined by measurement with the circuit of Fig. VII-17.

To solve for b_n in Eq. 11, we apply Eq. 12 to the mean square of that part of the output that comes from ϕ_n^*. Thus we obtain

$$\overline{\left\{\int h(\tau) b_n \left[\phi_n(t-\tau) + \sum_{j=1}^{n-1} a_{jn}\, \phi_j^*(t-\tau)\right] d\tau\right\}^2} = 1 \tag{15}$$

and solving for b_n, we get

$$b_n = \left[\overline{\left\{\int h(\tau) \left[\phi_n(t-\tau) + \sum_{j=1}^{n-1} a_{jn}\, \phi_j^*(t-\tau)\right] d\tau\right\}^2}\right]^{-1/2} \tag{16}$$

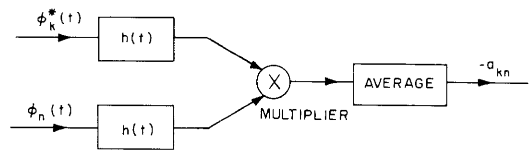

Fig. VII-17. Circuit for measurement of a_{kn}.

The average-value term in Eq. 16 can be measured directly in the manner described for a_{kn}.

Note that if the number of ϕ's is N, then the number of measurements necessary to form the set of ϕ^*'s is $(N+1)N/2$.

2. Nonlinear Fixed Network Without Memory

When the fixed network is a nonlinear, no-memory device, the necessary conditions given by Eqs. 5 and 6 are satisfied if the ϕ's are the gate functions that were used by Bose. If the set of ϕ's consists of gate functions, then at any instant one and only one ϕ is nonzero and that ϕ has unity output. It is this property of lack of coincidence in time of the gate-function ϕ's that satisfies Eqs. 5 and 6. In this case, the constants can be determined by independent measurements to satisfy any mean-function-of-error criterion in which the instantaneous function of error depends only on the instantaneous difference between r(t) and z(t). Mean-square error falls into this category, as does mean magnitude of error.

Example

As an example of the use of gate functions with a fixed nonlinear, no-memory network, consider the problem of determining a nonlinear filter to parallel the bandpass filter at the input of an FM receiver. See Fig. VII-18. The nonlinear, no-memory fixed network is the box in which the output of the nonlinear filter is added to the output

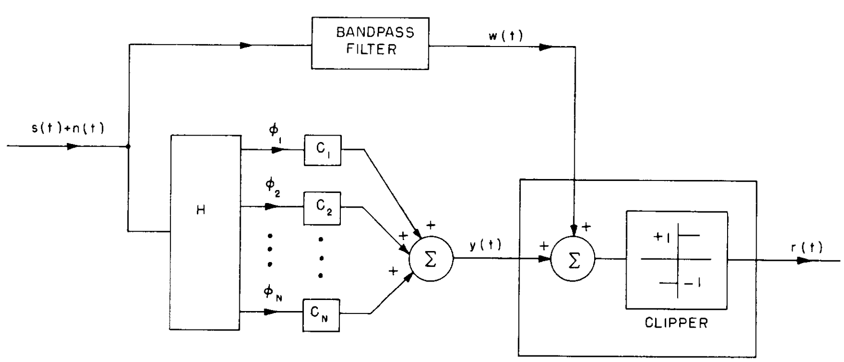

Fig. VII-18. FM filter.

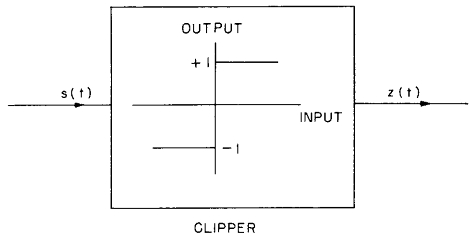

Fig. VII-19. Desired-signal generator.

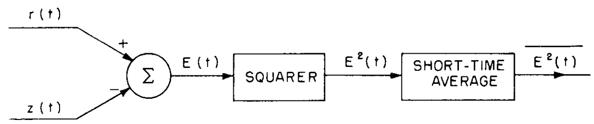

Fig. VII-20. Mean-square-error measurement circuit.

of the bandpass filter and then clipped. The input to the system is the signal s(t) plus the noise n(t). The desired signal output, z(t) – see Fig. VII-19 – is the clipped signal. The output of box H is a set of gate functions ϕ.

A circuit for measuring the mean-square error is given in Fig. VII-20. To determine C_i in the circuit of Fig. VII-18, vary C_i, keeping all the other C's fixed, until the minimum value of $\overline{E^2(t)}$ in Fig. VII-20 is measured.

3. Fixed Networks with Inverses

The methods of this section were suggested by Bose.

If the fixed network F has a realizable inverse F^{-1} which is such that

$$F(F^{-1}) = I \tag{17}$$

where I is the identity functional, then by preceding the fixed network by its inverse F^{-1}, we eliminate the effects of the fixed network.

If the fixed network F has a realizable inverse F_2^{-1} which is such that

$$F_2^{-1}(F) = I \tag{18}$$

then a new desired output $z^*(t)$ can be defined as

$$z^*(t) = F_2^{-1}(z) \tag{19}$$

The C's then can be adjusted to minimize the mean-square error given by

$$\overline{[E^*(t)]^2} = \overline{[y(t) - z^*(t)]^2} \tag{20}$$

The disadvantage of this latter procedure is that minimizing Eq. 20 does not, in general, minimize the mean-square error at the output of the fixed network.

D. A. Chesler

G. THEORY OF THE ANALYSIS OF NONLINEAR SYSTEMS

The present study has been completed. It was submitted as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science, Department of Electrical Engineering, M.I.T., January 13, 1958, and will also be presented as Technical Report 345.

M. B. Brilliant

H. A THEORY OF SIGNALS

This study has been completed. It was submitted as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science, Department of Electrical Engineering, M.I.T., January 31, 1958, and has been published as Technical Report 331.

R. E. Wernikoff

I. THEORY AND APPLICATION OF THE SEPARABLE CLASS OF RANDOM PROCESSES

The present study has been completed. It was submitted as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science, Department of Electrical Engineering, M.I.T., May 19, 1958, and will also be presented as Technical Report 343.

A. H. Nuttall