#3

DISCRETE REPRESENTATIONS OF RANDOM SIGNALS

KENNETH L. JORDAN, JR.

TECHNICAL REPORT 378

JULY 14, 1961

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory in which faculty members and graduate students from numerous academic departments conduct research.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U.S. Army (Signal Corps), the U.S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research) under Signal Corps Contract DA 36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

Technical Report 378 July 14, 1961

DISCRETE REPRESENTATIONS OF RANDOM SIGNALS

Kenneth L. Jordan, Jr.

Submitted to the Department of Electrical Engineering, M.I.T., August 22, 1960, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

The aim of this research is to present an investigation of the possibility of efficient, discrete representations of random signals. In many problems a conversion is necessary between a signal of continuous form and a signal of discrete form. This conversion should take place with small loss of information but still in as efficient a manner as possible.

Optimum representations are found for a finite time interval. The asymptotic behavior of the error in the stationary case is related to the spectrum of the process.

Optimal solutions can also be found when the representation is made in the presence of noise. These solutions are closely connected with the theory of optimum linear systems.

Some experimental results have been obtained by using these optimum representations.

TABLE OF CONTENTS

I. INTRODUCTION

1.1 THE PROBLEM OF SIGNAL REPRESENTATION

A signal represents the fluctuation with time of some quantity, such as voltage, temperature, or velocity, which contains information of some ultimate usefulness. It may be desired, for example, to transmit the information contained in this signal over a communication link to a digital computer on which mathematical operations will be performed. At some point in the system, the signal must be converted into a form acceptable to the computer, that is, a discrete or digital form. This conversion should take place with small loss of information and yet in as efficient a manner as possible. In other words, the digital form should retain only those attributes of the signal which are information-bearing.

The purpose of the research presented here has been to investigate the possibility of efficient, discrete representations of random signals.

Another example which involves the discrete representation of signals is the characterization of nonlinear systems described by Bose.[4] This involves the separation of the system into two sections, a linear section and a nonlinear, no-memory section. The linear section is the representation of the past of the input in terms of the set of Fourier coefficients of a Laguerre function expansion. The second section then consists of nonlinear, no-memory operations on these coefficients. Thus, the representation characterizes the memory of the nonlinear system. This idea originated with Wiener.

The study presented in this report actually originated from a suggestion by Professor Amar G. Bose in connection with this characterization of nonlinear systems. He suggested that since in practice we shall only use a finite number of Fourier coefficients to represent the past of a signal, perhaps some set of functions other than Laguerre functions might result in a better representation. We have been able to solve this problem with respect to a weighted mean-square error or even a more general criterion. The problem of finding the best representation with respect to the operation of the nonlinear system as a whole has not been solved.

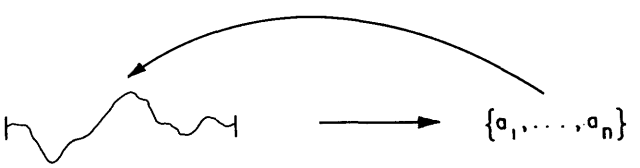


Fig. 1. Discrete representation of a random function.

The problem of discrete representation as it is considered in this report is illustrated in Fig. 1. A set of numbers that are random variables is derived from a random process x(t) and represents that process in some way. We must then be able to use the

information contained in the set of random variables to return to a reasonable approximation of the process x(t). The fidelity of the representation is then measured by finding how close we come to x(t) with respect to some criterion.

1.2 THE HISTORY OF THE PROBLEM

The problem of discrete representation of signals has been considered by many authors, including Shannon,[28] Balakrishnan,[1] and Karhunen.[19] Shannon and Balakrishnan considered sampling representations, and Karhunen has done considerable work on series representations. To our knowledge, the only author who has done a great deal of thinking along the lines of efficient representations is Huggins.[11] He considered exponential representations of signals which are especially useful when dealing with speech waveforms.

II. BASIC CONCEPTS

We shall now briefly present some of the fundamental ideas that will form the basis of the following work. The first three sections will cover function spaces and linear methods. A theorem that will be used several times is presented in the fourth section. The fifth section will discuss random processes and some methods of decomposition. This section is intended to give a résumé, and the only part that is original with the author is a slight extension of Fan's theorem given in section 2.4.

2.1 FUNCTION SPACE

A useful concept in the study of linear transformations and approximations of square integrable functions is the analogy of functions with vectors (function space). As we can express any vector $\underline{v}$ in a finite dimensional vector space as a linear combination of a set of basis vectors $\{\underline{\phi}_i\}$

$$\underline{v} = \sum_{i=1}^{n} v_i \underline{\phi}_i \tag{1}$$

so can we express any square integrable function defined on an interval Ω as an infinite linear combination of a set of basis functions

$$f(t) = \sum_{i=1}^{\infty} a_i \phi_i(t) \tag{2}$$

The analogy is not complete, however, since, in general, the equality sign in Eq. (2) does not necessarily hold for all $t \in \Omega$. If the a_i are chosen in a certain way, the equality can always be interpreted in the sense that

$$\lim_{n\to\infty} \int_{\Omega} \left[f(t) - \sum_{i=1}^{n} a_i \phi_i(t) \right]^2 dt = 0 \tag{3}$$

To be precise we should say that the series converges in the mean to f(t) or

$$f(t) = \text{l.i.m.}_{n\to\infty} \sum_{i=1}^{n} a_i \phi_i(t)$$

where l.i.m. stands for "limit in the mean." Moreover, if it can be shown that the series converges uniformly, then the equality will be good for all $t \in \Omega$, that is

$$f(t) = \lim_{n\to\infty} \sum_{i=1}^{n} a_i \phi_i(t)$$

(If, for any strip $(f(t)+\epsilon, f(t)-\epsilon)$ for $t \in \Omega$, the approximation $\sum_{i=1}^{n} a_i \phi_i(t)$ lies within the strip for n large enough, the series converges uniformly. For a discussion of uniform and nonuniform convergence, see Courant.[32])

If the set $\{\phi_i(t)\}$ is orthonormal

$$\int_\Omega \phi_i(t)\, \phi_j(t)\, dt = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and complete; that is,

$$\int_\Omega \phi_i(t)\, f(t)\, dt = 0 \qquad \text{all } i = 1, 2, \ldots$$

if and only if $\int_\Omega f^2(t)\, dt = 0$, then the coefficients a_i in Eq. (2) can be given by

$$a_i = \int_\Omega f(t)\, \phi_i(t)\, dt$$

and the limit (3) holds.

Uniform convergence is certainly a stronger condition than convergence in the mean, and in most cases is much more difficult to establish. If we are interested in approximating the whole function, in most engineering situations we shall be satisfied with convergence in the mean, since Eq. (3) states that the energy in the error can be made as small as is desired. If, on the other hand, we are interested in approximating the function at a single time instant, convergence in the mean does not insure convergence for that time instant and we shall be more interested in establishing uniform convergence.

Another useful concept that stems from the analogy is that of length. Ordinary Euclidian length as defined in a finite dimensional vector space is

$$|\underline{v}| = \left[\sum_{i=1}^{n} v_i^2 \right]^{1/2}$$

and in function space it can be defined as

$$|f(t)| = \left[\int_\Omega f^2(t)\, dt \right]^{1/2}$$

It can be shown that both these definitions satisfy the three conditions that length in ordinary Euclidean three-dimensional space satisfies:

(i) $|\underline{v}| = 0$, if and only if $\underline{v} = 0$.

(ii) $|c\underline{v}| = c|\underline{v}|$

(iii) $|\underline{v}+\underline{w}| \leq |\underline{v}| + |\underline{w}|$

The first states that the length of a vector is zero if and only if all of its components are zero; the second is clear; and the third is another way of saying that the shortest distance between two points is a straight line.

There are other useful definitions of length which satisfy these conditions, for example

$$|f(t)| = \left[\int_\Omega W(t)\, f^2(t)\, dt\right]^{1/2}$$

where $W(t) > 0$. We shall call any such definition a "norm," and we shall denote a norm by $\|f(t)\|$ or $\|f\|$.

In other sections we shall use the norm as a measure of the characteristic differences between functions. Actually, it will not be necessary to restrict ourselves to a measure that satisfies the conditions for a norm, and we do so only to retain the geometric picture.

In vector space we also have the inner product of two vectors (We use the bracket notation $\langle \underline{v}, \underline{w} \rangle$ to denote the inner product.)

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i w_i$$

and its analogous definition in function space is

$$\langle f, g \rangle = \int_\Omega f(t)\, g(t)\, dt$$

An important concept is the transformation or "operator." In vector space, an operator L is an operation which when applied to any vector $\underline{v}$ gives another vector

$$\underline{w} = L[\underline{v}]$$

It is a linear operator when

$$L[a_1\underline{v}_1 + a_2\underline{v}_2] = a_1 L[\underline{v}_1] + a_2 L[\underline{v}_2]$$

for any two vectors $\underline{v}_1$ and $\underline{v}_2$. Any linear operation in a finite dimensional space can be expressed as

$$w_i = \sum_{j=1}^{n} a_{ij} v_j \qquad i = 1, \ldots, n$$

which is the matrix multiplication

$$w_i] = [a_{ij}]\, v_i]$$

The same definition holds in function space, and we have

$$g(t) = L[f(t)]$$

A special case of a linear operator is the integral operator

$$g(s) = \int_\Omega K(s,t)\, f(t)\, dt$$

where $K(s,t)$ is called the "kernel" of the operator.

A "functional" is an operation which when applied to a vector gives a number; that is,

$$c = T[\underline{v}]$$

and a linear functional obeys the law

$$T[a_1\underline{v}_1 + a_2\underline{v}_2] = a_1 T[\underline{v}_1] + a_2 T[\underline{v}_2]$$

For a function space, we have $c = T[f(t)]$. The norm and an inner product with a particular function are functionals. In fact, it can be shown that a particular class of linear functionals can always be represented as an inner product, that is,

$$T[f(t)] = \int_\Omega f(t)\, g(t)\, dt$$

for any $f(t)$. (These are the bounded or continuous linear functionals. See Friedman.[33])

2.2 INTEGRAL EQUATIONS

There are two types of integral equation that will be considered in the following work. These are

$$\int_\Omega K(s,t)\, \phi(t)\, dt = \lambda\phi(s) \qquad s \in \Omega \tag{4}$$

where the unknowns are $\phi(t)$ and λ, and

$$\int_\Omega K(s,t)\, g(t)\, dt = f(s) \qquad s \in \Omega \tag{5}$$

where the unknown is $g(t)$.

The solutions of the integral (4) have many properties and we shall list a number of these that will prove useful later. We shall assume that

$$\int_\Omega \int_\Omega |K(s,t)|^2\, ds\, dt < \infty$$

and that the kernel is real and symmetric, $K(s,t) = K(t,s)$. The solutions $\phi_i(t)$ of (4) are called the eigenfunctions of $K(s,t)$, and the corresponding set $\{\lambda_i\}$ is the set of eigenvalues or the "spectrum." We have the following properties:

(a) The spectrum is discrete; that is, the set of solutions is a countable set. (A proof has been given by Courant and Hilbert.[34])

(b) Any two eigenfunctions corresponding to distinct eigenvalues are orthogonal. If there are n linearly independent solutions corresponding to an eigenvalue λ_i, it is said that λ_i has multiplicity n. These n solutions can be orthogonalized by the Gram-Schmidt procedure, and in the following discussion we shall assume that this has been done. (A proof has been given by Petrovskii.[35])

(c) If the kernel $K(s,t)$ is positive definite; that is,

$$\int_\Omega \int_\Omega K(s,t)\, f(s)\, f(t)\, ds\, dt > 0$$

for $f(t) \neq 0$, then the set of eigenfunctions is complete. (A proof has been given by Smithies.[36])

(d) The kernel $K(s,t)$ may be expressed as the series of eigenfunctions

$$K(s,t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s)\, \phi_i(t) \tag{6}$$

which is convergent in the mean. (A proof has been given by Petrovskii.[37])

(e) If $K(s,t)$ is non-negative definite; that is,

$$\int_\Omega \int_\Omega K(s,t)\, f(s)\, f(t)\, ds\, dt \geq 0$$

for any $f(t)$, then the series (6) converges absolutely and uniformly (Mercer's theorem). (A proof has been given by Petrovskii.[38])

(f) If

$$f(s) = \int_\Omega K(s,t)\, g(t)\, dt$$

where $g(t)$ is square integrable, then $f(s)$ can be expanded in an absolutely and uniformly convergent series of the eigenfunctions of $K(s,t)$ (Hilbert-Schmidt theorem). (A proof has been given by Petrovskii.[39])

(g) A useful method for characterizing the eigenvalues and eigenfunctions of a kernel utilizes the extremal property of the eigenvalues. The quadratic form

$$\int_\Omega \int_\Omega K(s,t)\, f(s)\, f(t)\, ds\, dt$$

where f(t) varies under the conditions

$$\int_\Omega f^2(s)\, ds = 1$$

$$\int_\Omega f(s)\, \gamma_i(s)\, ds = 0 \qquad i = 1, 2, \ldots, n-1$$

where the $\gamma_i(t)$ are the eigenfunctions of $K(s,t)$, is maximized by the choice $f(t) = \gamma_n(t)$, and the maximum is λ_n. There exists also a minimax characterization that does not require the knowledge of the lower-order eigenfunctions. (A proof has been given by Smithies.[40])

We shall adopt the convention that zero is a possible eigenvalue so that every set of eigenfunctions will be considered complete.

By Picard's theorem (see Courant and Hilbert[41]), Eq. (5) has a square integrable solution if and only if the series

$$\sum_{i=1}^{\infty} \frac{\left| \int_\Omega f(t)\, \gamma_i(t)\, dt \right|^2}{\lambda_i^2}$$

converges. The solution is then

$$g(t) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \gamma_i(t) \int_\Omega f(t)\, \gamma_i(t)\, dt \qquad t \in \Omega$$

2.3 THE SPECTRAL REPRESENTATION OF A LINEAR OPERATOR

A useful tool in the theory of linear operators is the spectral representation. (An interesting discussion of this topic is given by Friedman.[42]) Let us consider the operator equation

$$L[\phi(t)] = \lambda \phi(t) \tag{7}$$

where the linear operator L is "self-adjoint"; that is,

$$\langle f, L[g] \rangle = \langle L[f], g \rangle$$

An example of such an operator equation is the integral Eq. (4) where the kernel is assumed symmetric. It is self-adjoint, since

$$\langle f, L[g] \rangle = \int_\Omega f(s) \left\{ \int_\Omega K(s,t)\, g(t)\, dt \right\} ds$$

$$= \int_\Omega \left\{ \int_\Omega K(t,s)\, f(s)\, ds \right\} g(t)\, dt$$

$$= \langle L[f], g \rangle$$

The solutions of Eq. (7) are the eigenvalues and eigenfunctions of L, and the set of eigenvalues $\{\lambda_i\}$ is called the spectrum.

We shall assume that Eq. (7) has a countable number of solutions; that is, $\{\lambda_i\}$ is a discrete spectrum. It can be shown that any two eigenfunctions corresponding to distinct eigenvalues are orthogonal[43]; therefore, if the set of eigenfunctions is complete, we can assume that it is a complete, orthonormal set. If $\{\gamma_i(t)\}$ is such a set of eigenfunctions, then any square integrable function f(t) may be expanded as follows:

$$f(t) = \sum_{i=1}^{\infty} f_i \gamma_i(t) \tag{8}$$

If we apply L, we get

$$L[f(t)] = \sum_{i=1}^{\infty} f_i \lambda_i \gamma_i(t) \tag{9}$$

The representation of f(t) and L[f(t)] in Eqs. (8) and (9) is called the "spectral representation" of L. It is seen that the set of eigenvalues and eigenfunctions completely characterizes L.

If we want to solve the equation L[f(t)] = g(t) for f(t), then we use the spectral representation and we get

$$f(t) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} g_i \gamma_i(t)$$

It is then seen that the eigenvalues $1/\lambda_i$ and eigenfunctions $\gamma_i(t)$ characterize the inverse L^{-1} of L. For example, if we have an integral operator with a kernel $K(s,t) = \sum_{i=1}^{\infty} \lambda_i \gamma_i(s)\, \gamma_i(t)$, then the inverse operator is characterized by $1/\lambda_i$ and $\gamma_i(t)$ and we can write

$$K^{-1}(s,t) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \gamma_i(s)\, \gamma_i(t)$$

where $K^{-1}(s,t)$ is the inverse kernel; this makes sense only if the series converges.

It is also interesting to note that if we define an operator L^n to be the operation L taken n times; that is,

$$L^n[f(t)] = L[L[\ldots L[f(t)]\ldots]$$

then the spectrum of L^n is $\left\{\lambda_i^n\right\}$, where $\{\lambda_i\}$ is the spectrum of L, and the eigenfunctions are identical.

It must be pointed out that the term "spectrum" as used here is not to be confused with the use of the word in connection with the frequency spectrum or power density spectrum of a random process. There is a close relation, however, between the spectrum of a linear operator and the system function of a linear, time-invariant system. Consider the operation

$$y(t) = \int_{-\infty}^{\infty} h(t-s)\, x(s)\, ds$$

where $h(t) = h(-t)$. This is a time-invariant operation with a symmetrical kernel. The equation

$$\int_{-\infty}^{\infty} h(t-s)\, \phi(s)\, ds = \lambda\phi(t)$$

is satisfied by any function of the form

$$\phi_f(t) = e^{j2\pi ft}$$

where

$$\lambda_f = H(f) = \int_{-\infty}^{\infty} h(t)\, e^{-j2\pi ft}\, dt$$

Thus, we have a continuum of eigenvalues and eigenfunctions and H(f) is the continuous spectrum, or, as it is known in linear system theory, the "system function." This is a useful representation because if we cascade two time-invariant systems with system functions $H_1(f)$ and $H_2(f)$, the system function of the resultant is $H_1(f)\, H_2(f)$. A similar relation occurs for the spectra of linear operators with the same eigenvalues. If we cascade two linear operators with spectra $\left\{\lambda_i^{(1)}\right\}$ and $\left\{\lambda_i^{(2)}\right\}$, the spectrum of the resultant linear operator is $\left\{\lambda_i^{(1)}\lambda_i^{(2)}\right\}$.

2.4 A USEFUL THEOREM

We now consider a theorem that is a slight extension of a theorem of Fan.[44,45] Suppose that L is a self-adjoint operator with a discrete spectrum and suppose that it has a maximum (or minimum) eigenvalue. The eigenvalues and eigenfunctions of L are $\lambda_1, \lambda_2, \ldots$ and $\gamma_1(t), \gamma_2(t), \ldots$ arranged in descending (ascending) order. We then

state the following theorem which is proved in Appendix A.

THEOREM I. The sum

$$\sum_{i=1}^{n} c_i \left\langle \phi_i, L[\phi_i] \right\rangle$$

where $c_1 \geq c_2 \geq \ldots \geq c_n$, is maximized (minimized) with respect to the orthonormal set of functions $\{\phi_i(t)\}$ by the choice

$$\phi_i(t) = \gamma_i(t) \qquad i = 1, 2, \ldots, n$$

and this maximum (minimum) value is $\sum_{i=1}^{n} c_i \lambda_i$. It is useful to state the corollary for the integral operator $L[f(t)] = \int_{\Omega} K(s,t)\, f(t)\, dt$.

COROLLARY. The sum

$$\sum_{i=1}^{n} \int_{\Omega} \int_{\Omega} K(s,t)\, \phi_i(s)\, \phi_i(t)\, ds\, dt$$

is maximized with respect to the orthonormal set of functions $\{\phi_i(t)\}$ by the choice

$$\phi_i(t) = \gamma_i(t) \qquad i = 1, 2, \ldots, n$$

and the maximum value is $\sum_{i=1}^{n} \lambda_i$, where the λ_i and $\gamma_i(t)$ are the eigenvalues and eigenfunctions of $K(s,t)$.

2.5 RANDOM PROCESSES AND THEIR DECOMPOSITION

For a random process $x(t)$, we shall generally consider as relevant statistical properties the first two moments

$$m(t) = E[x(t)]$$

$$r(s,t) = E[(x(s)-m(s))(x(t)-m(t))]$$

Here, $m(t)$ is the mean-value function, and $r(s,t)$ is the autocovariance function. We also have the autocorrelation function $R(s,t) = E[x(s)\, x(t)]$ which is identical with the autocovariance function if the mean is zero. For stationary processes, $R(s,t) = R(s-t)$ and we have the Wiener-Khinchin theorem

$$R(t) = \int_{-\infty}^{\infty} S(f)\, e^{j2\pi ft}\, df$$

and its inverse

$$S(f) = \int_{-\infty}^{\infty} R(t)\, e^{-j2\pi ft}\, dt$$

where S(f) is the power density spectrum of the process x(t).

Much of the application to random processes of the linear methods of function space is due to Karhunen.[19,20] The Karhunen-Loeve expansion theorem[46] states that a random process in an interval of time Ω may be written as an orthonormal series with uncorrelated coefficients. Suppose that x(t) has mean m(t) and autocovariance r(s,t). The autocovariance is non-negative definite and by considering the integral equation

$$\int_{\Omega} r(s,t)\, \gamma_i(t)\, dt = \lambda_i \gamma_i(s) \qquad s \in \Omega$$

we get the expansion

$$x(t) = m(t) + \sum_{i=1}^{\infty} a_i \gamma_i(t) \qquad t \in \Omega \tag{10}$$

for which

$$E[a_i a_j] = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}$$

where $a_i = \int_{\Omega} (x(t)-m(t))\, \gamma_i(t)\, dt$ for i = 1, 2, ... Moreover, the representation (10) converges in the mean for every t. (This is convergence in the mean for random variables, which is not to be confused with convergence in the mean for functions. A sequence of random variables x_n converges in the mean to the random variable x if and only if $\lim_{n\to\infty} E\left[(x-x_n)^2\right] = 0$.) This is a direct consequence of Mercer's theorem, since

$$E\left[\left[x(t) - m(t) - \sum_{i=1}^{n} a_i \gamma_i(t)\right]^2\right] = r(t,t) - 2\sum_{i=1}^{n} \gamma_i(t) \int_{\Omega} r(t,s)\, \gamma_i(s)\, ds + \sum_{i=1}^{n} \lambda_i \gamma_i^2(t)$$

$$= r(t,t) - \sum_{i=1}^{n} \lambda_i \gamma_i^2(t)$$

By Mercer's theorem, $\lim_{n\to\infty} \sum_{i=1}^{n} \lambda_i \gamma_i^2(t) = r(t,t)$, and therefore

$$\lim_{n\to\infty} E\left[\left[x(t) - m(t) - \sum_{i=1}^{n} a_i \gamma_i(t)\right]^2\right] = 0$$

Karhunen has given another representation theorem which is the infinite analog of the Karhunen-Loeve representation. Let x(t) be a process with zero mean and

autocorrelation function $R(s,t)$, and suppose that $R(s,t)$ is expressible in the form of the Stieltjes integral[47]

$$R(s,t) = \int_{-\infty}^{\infty} f(s,u)\, f(t,u)\, d\sigma(u)$$

where $\sigma(u)$ is a nondecreasing positive function of u. There exists, then, an orthogonal process $Z(s)$ so that

$$x(t) = \int_{-\infty}^{\infty} f(t,s)\, dZ(s)$$

where $E[Z^2(s)] = \sigma(s)$. (A process is orthogonal if for any two disjoint intervals (u_1, u_2) and (u_3, u_4), $E[(Z(u_2)-Z(u_1))(Z(u_4)-Z(u_3))] = 0$.) If, in particular, the process $x(t)$ is stationary, then, from the Wiener-Khinchin theorem, we have

$$R(s-t) = \int_{-\infty}^{\infty} e^{j2\pi f(s-t)}\, dF(f)$$

in the form of a Stieltjes integral, so that we obtain the "spectral decomposition" of the stationary process,

$$x(t) = \int_{-\infty}^{\infty} e^{j2\pi ft}\, dZ(f)$$

which is due to Cramér.

III. THE THEORY OF DISCRETE REPRESENTATIONS

3.1 GENERAL FORMULATION

An important aspect of any investigation is the formulation of the general problem. It gives the investigator a broad perspective so that he may discern the relation of those questions that have been answered to the more general problem. It also aids in giving insight into the choice of lines of further investigation.

In the general formulation of the problem of discrete representation, we must be able to answer three questions with respect to any particular representation:

(a) How is the discrete representation derived from the random process x(t)?

(b) In what way does it represent the process?

(c) How well is the process represented?

To answer these questions it is necessary to give some definitions:

(a) We shall define a set of functionals $\{T_i\}$ by which the random variables $\{a_i\}$ are derived from x(t), that is, $a_i = T_i[x(t)]$.

(b) For transforming the set $\{a_i\}$ into a function z(t) which in some sense approximates x(t), we need to define an approximation function F for which $z(t) = F(t, a_1, \ldots, a_n)$.

(c) We must state in what sense z(t) approximates x(t) by introducing a norm on the error, $\|e(t)\| = \|x(t)-z(t)\|$. (In general, it would not be necessary to restrict ourselves to a norm here; however, it is convenient for our purposes.) This norm shall comprise the criteria for the relative importance of the characteristics of x(t).

(d) We must utilize a statistical property of $\|e(t)\|$ to obtain a fidelity measure across the ensemble of x(t). In this report we use $\theta = E[\|e(t)\|^2]$, although others could be defined. (For example, $P[\|e(t)\| \geq k]$. It may be well to point out, however, that the choice of the expected value is not arbitrary but made from the standpoint of analytical expediency.) We shall sometimes refer to θ as "the error."

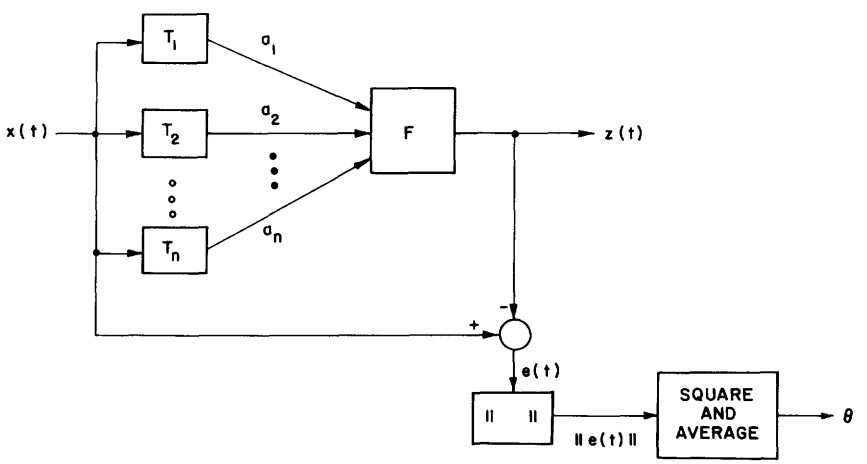


Fig. 2. The process of fidelity measurement of a discrete representation.

The process of fidelity measurement of a discrete representation would then be as shown by the block diagram in Fig. 2.

We are now in a position to state the fundamental problem in the study of the discrete representation of random signals. We must so determine the set $\{T_i\}$ and F that

$$\theta = E[\,\|x(t)-F(t, a_1, \ldots, a_n)\|^2] \tag{11}$$

shall be a minimum. We shall denote this minimum value by θ^* and the $\{T_i\}$ and F for which it is attained by $\{T_i^*\}$ and F^*. In many cases the requirements of the problem may force the restriction of $\{T_i\}$ and F to certain classes, in which case we would perform the minimization discussed above with the proper constraints.

It is certain that the solution of this problem, in general, would be a formidable task. We shall be dealing largely with those cases in which $\{T_i\}$ and F are linear and the norm is the square root of a quadratic expression. This is convenient because the minimization of Eq. (11) then simply requires the solution of linear equations.

3.2 LINEAR REPRESENTATIONS IN A FINITE TIME INTERVAL

We shall now consider a random process x(t) to be represented in a finite interval of time. We shall assume that (a) the approximation function F is constrained to be of no higher than first degree in the variables $a_1, \ldots, a_n$, and (b) the norm is $\|f(t)\| = \left[\int_\Omega |f(t)|^2\, dt\right]^{1/2}$, where the interval of integration, Ω, is the region of t over which the process is to be represented.

Considering t as a parameter, we see that $F(t, a_1, \ldots, a_n)$ may be written

$$F(t, a_1, \ldots, a_n) = c(t) + \sum_{i=1}^{n} a_i \phi_i(t)$$

We then want to minimize

$$\theta = E\left[\int_\Omega \left| x(t) - c(t) - \sum_{i=1}^{n} a_i \phi_i(t)\right|^2 dt\right] \tag{12}$$

The minimization will be performed, first, with respect to the functionals $\{T_i\}$ while F is assumed to be arbitrary (subject to the constraint) but fixed. There is no restriction in assuming that the set of functions $\{\phi_i(t)\}$ is an orthonormal set over the interval Ω, for if it is not, we can put F into such a form by performing a Gram-Schmidt orthogonalization.[48]

We have then $\int_\Omega \phi_i(t)\, \phi_j(t)\, dt = \delta_{ij}$, where δ_{ij} is the Kronecker delta.

It follows from the minimal property of Fourier coefficients[49] that the quantity in brackets in Eq. (12) is minimized for each x(t) by the choice

$$a_i = T_i[x(t)] = \int_\Omega [x(t)-c(t)]\, \phi_i(t)\, dt \qquad i = 1, \ldots, n$$

over all possible sets $\{T_i\}$. Likewise, it follows that its expected value, θ, must be minimized. Setting $y(t) = x(t) - c(t)$, we see that the minimized expression is

$$\theta = E\left[\int_\Omega \left| y(t) - \sum_{i=1}^{n} \phi_i(t) \int_\Omega y(s)\, \phi_i(s)\, ds \right|^2 dt\right]$$

$$= \int_\Omega R_y(t,t)\, dt - \sum_{i=1}^{n} \int_\Omega \int_\Omega R_y(s,t)\, \phi_i(s)\, \phi_i(t)\, ds\, dt$$

By the corollary of Theorem I, we know that

$$\sum_{i=1}^{n} \int_\Omega \int_\Omega R_y(s,t)\, \phi_i(s)\, \phi_i(t)\, ds\, dt \leq \sum_{i=1}^{n} \int_\Omega \int_\Omega R_y(s,t)\, \gamma_i(s)\, \gamma_i(t)\, ds\, dt = \sum_{i=1}^{n} \lambda_i$$

where the λ_i and the $\gamma_i(t)$ are the eigenvalues and eigenfunctions of the kernel $R_y(s,t)$. θ is then minimized with respect to the $\phi_i(t)$ by the choice $\phi_i(t) = \gamma_i(t)$. The error now is

$$\theta = \int_\Omega R_y(t,t)\, dt - \sum_{i=1}^{n} \lambda_i$$

From Mercer's Theorem (see section 2.2), we have

$$R_y(s,t) = \sum_{i=1}^{\infty} \lambda_i \gamma_i(s)\, \gamma_i(t)$$

so that

$$\int_\Omega R_y(t,t)\, dt = \sum_{i=1}^{\infty} \lambda_i$$

and therefore

$$= \sum_{i=1}^{\infty} \lambda_i - \sum_{i=1}^{n} \lambda_i = \sum_{i=n+1}^{\infty} \lambda_i$$

We now assert that each eigenvalue is minimized by choosing $c(t) = m_x(t) = E[x(t)]$. We have for each eigenvalue

$$\lambda_i = \int_\Omega \int_\Omega R_y(s,t)\, \gamma_i(s)\, \gamma_i(t)\, ds\, dt$$

$$= \int_\Omega \int_\Omega E[x(s)x(t)-x(s)c(t)-c(s)x(t)+c(s)c(t)]\, \gamma_i(s)\, \gamma_i(t)\, ds\, dt$$

$$= \int_\Omega \int_\Omega R_x(s,t)\, \gamma_i(s)\, \gamma_i(t)\, ds\, dt - 2 \int_\Omega m_x(s)\, \gamma_i(s)\, ds \int_\Omega c(t)\, \gamma_i(t)\, dt$$

$$+ \left[\int_\Omega c(s)\, \gamma_i(s)\, ds\right]^2 \qquad (13)$$

Now, since

$$\left| \int_\Omega c(s)\, \gamma_i(s)\, ds - \int_\Omega m_x(s)\, \gamma_i(s)\, ds \right|^2 \geq 0$$

We have

$$\left| \int_\Omega c(s)\, \gamma_i(s)\, ds \right|^2 - 2 \int_\Omega m_x(s)\, \gamma_i(s)\, ds \int_\Omega c(t)\, \gamma_i(t)\, dt \geq - \left| \int_\Omega m_x(s)\, \gamma_i(s)\, ds \right|^2$$

Here, the equality sign holds for

$$\int_\Omega m_x(s)\, \gamma_i(s)\, ds = \int_\Omega c(t)\, \gamma_i(t)\, dt$$

After applying this inequality to Eq. (13) we find that λ_i is minimum for

$$\int_\Omega m_x(s)\, \gamma_i(s)\, ds = \int_\Omega c(t)\, \gamma_i(t)\, dt$$

and since we want this to hold for all i, we have $c(t) = m_x(t)$.

So, we finally see that if we have a random process x(t) with mean $m_x(t)$ and covariance function $r_x(s,t) = E[\{x(s)-m_x(s)\}\{x(t)-m_x(t)\}]$, then θ is minimized for

$$F^*(t, a_1, \ldots, a_n) = m_x(t) + \sum_{i=1}^{n} a_i \gamma_i(t)$$

The $\gamma_i(t)$ are the solutions of

$$\int_\Omega r_x(s,t)\, \gamma_i(t)\, dt = \lambda_i \gamma_i(s) \qquad s \in \Omega$$

arranged in the order $\lambda_1 \geq \lambda_2 \geq \ldots$, and

$$a_i^* = \int_\Omega x(t)\, \gamma_i(t)\, dt - \int_\Omega m_x(t)\, \gamma_i(t)\, dt$$

The minimum error is, then,

$$\theta^* = \int_\Omega r_x(t,t)\, dt - \sum_{i=1}^{n} \lambda_i = \sum_{i=n+1}^{\infty} \lambda_i \tag{14}$$

This solution is identical to the Karhunen-Loeve expansion of a process in an orthonormal series with uncorrelated coefficients which was described in section 2.4. This was first proved by Koschmann,[21] and has since been discussed by several other authors.[13, 5, 22]

We have assumed that x(t) has a nonzero mean. In the solution, however, the mean is subtracted from the process and for the reconstruction it is added in again. In the rest of this report we shall consider, for the most part, zero-mean processes, for if they are not, we can subtract the mean.

3.3 A GEOMETRICAL DESCRIPTION

A useful geometric picture can be obtained by considering a random process in a finite time interval as a random vector in an infinite dimensional vector space. This geometric picture will be used in this section in order to gain understanding of the result of section 3.2, but we shall confine ourselves to a finite m-dimensional vector space. The process x(t) will then be representable as a finite linear combination of some orthonormal set of basis functions $\{\psi_i(t)\}$; that is,

$$x(t) = \sum_{i=1}^{m} x_i \psi_i(t)$$

where the x_i are the random coordinates of x(t). We see, then, that x(t) is equivalent to the random vector $\underline{x} = \{x_1, \ldots, x_m\}$.

We shall assume that x(t) has mean zero and correlation function $R_x(s,t)$. The random vector $\underline{x}$ then has mean zero and covariance matrix of elements $r_{ij} = E[x_i x_j]$, with

$$R_x(s,t) = \sum_{i=1}^{m} \sum_{j=1}^{m} r_{ij} \psi_i(s)\, \psi_j(t)$$

Our object is to represent $\underline{x}$ by a set of n random variables $\{a_1, \ldots, a_n\}$, with $n < m$. Using arguments similar to those of section 3.2, we see that we want to find the random vector $\underline{z} = \underline{c} + \sum_{i=1}^{n} a_i \underline{\phi}_i$ which minimizes $\theta = E[|\underline{x}-\underline{z}|^2]$. Since $\underline{x}$ has zero mean, we shall assume that $\underline{c} = \underline{0}$. Then, $\underline{z}$ is a random vector confined to an n-dimensional hyperplane through the origin. Since the set $\{\underline{\phi}_i\}$ determines the orientation of this plane, there is no restriction in assuming that it is orthonormal; that is, $\langle \underline{\phi}_i, \underline{\phi}_j \rangle = \delta_{ij}$. If we

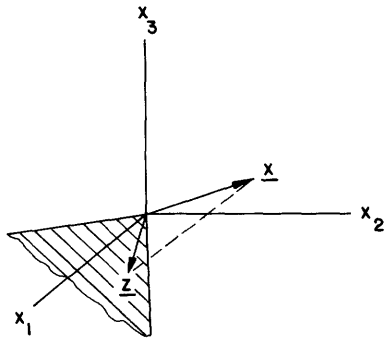


Fig. 3. The best approximation of a random vector.

are given a particular orientation for the plane, that is, a particular set $\{\underline{\phi}_i\}$, and a particular outcome of $\underline{x}$, then it is clear that the best $\underline{z}$ is the projection of $\underline{x}$ onto the plane, as shown in Fig. 3. That is, $\underline{z} = \sum_{i=1}^{n} \langle \underline{x}, \underline{\phi}_i \rangle \underline{\phi}_i$, so that $a_i = \langle \underline{x}, \underline{\phi}_i \rangle$, (i=1,...,n). This is related to the minimal property of Fourier coefficients, as mentioned in section 3.2. The error, θ, then becomes

$$\theta = E[|\underline{x}-\underline{z}|^2]$$

$$= E\left[\left\langle \underline{x} - \sum_{i=1}^{n} \langle \underline{x}, \underline{\phi}_i \rangle \underline{\phi}_i,\ \underline{x} - \sum_{i=1}^{n} \langle \underline{x}, \underline{\phi}_i \rangle \underline{\phi}_i \right\rangle\right]$$

$$= E[|\underline{x}|^2] - E\left[\sum_{i=1}^{n} \langle \underline{x}, \underline{\phi}_i \rangle^2\right] \qquad (15)$$

Now, we must find the orientation of the hyperplane which minimizes θ. From Eq. (15), we see that this is equivalent to finding the orientation that maximizes the average of the squared length of the projection of $\underline{x}$. We have for the inner product

$$\langle \underline{x}, \underline{\phi}_i \rangle = \sum_{j=1}^{m} x_j \phi_{ij}$$

where $\underline{\phi}_i = \{\phi_{i1}, \ldots \phi_{im}\}$. Then

$$\theta = E[|\underline{x}|^2] - E\left[\sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} x_j \phi_{ij} \right\}^2 \right]$$

$$= E[|\underline{x}|^2] - \sum_{i=1}^{n} \left[\sum_{j=1}^{m} \sum_{k=1}^{m} r_{jk} \phi_{ij} \phi_{ik} \right]$$

The quantity in brackets is a quadratic form

$$\mathcal{F}[\phi_{i1},\ldots,\phi_{im}] \equiv \sum_{j=1}^{m}\sum_{k=1}^{m} r_{jk}\phi_{ij}\phi_{ik}$$

so that we must maximize

$$\sum_{i=1}^{n} \mathcal{F}[\phi_{i1},\ldots,\phi_{im}]$$

where $\{\underline{\phi_i}\}$ is constrained to be an orthonormal set.

Suppose that n = 1, than we must maximize $\mathcal{F}[\phi_{11},\ldots,\phi_{1m}]$ subject to the condition $|\underline{\phi_1}| = 1$. By the maximum property of the eigenvalues mentioned in section 2.2 we see that

$$\max_{\underline{\phi_1}} \mathcal{F}[\underline{\phi_1}] = \mathcal{F}[\underline{\gamma_1}] = \lambda_1$$

$$|\underline{\phi_1}| = 1$$

where λ_1 is the largest eigenvalue of the positive definite matrix $[r_{ij}]$, and $\underline{\gamma_1}$ is the corresponding eigenvector. So we have the solution for n = 1. The surface generated

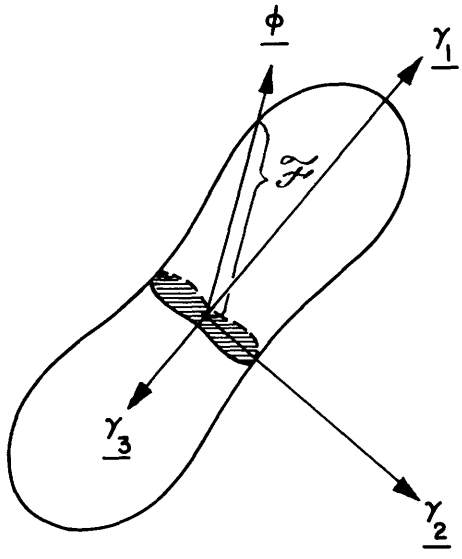


Fig. 4. The surface generated by the quadratic form.

by $\mathcal{F}$, by allowing $\underline{\phi_1}$ to take on all possible orientations, would be similar to that shown in Fig. 4 for m = 3. This surface has the property that $\sum_{i=1}^{m} \mathcal{F}[\underline{\phi_i}]$ is invariant with respect to the set $\{\underline{\phi_i}\}$ and is equal to $\sum_{i=1}^{m} \lambda_i$. This must be so, since if all m dimensions

are used in the approximation, the error must be zero.

By the maximum property of the eigenvalues, we also have

$$\max_{\underline{\phi}_i} \mathscr{F}[\underline{\phi}_i] = \mathscr{F}[\underline{\gamma}_i] = \lambda_i$$

$$\langle \underline{\phi}_i, \underline{\gamma}_j \rangle = 1 \qquad j = 1, \ldots, i-1$$

So, from this property and by observing Fig. 4 we might expect that

$$\max_{\{\underline{\phi}_i\}} \sum_{i=1}^{n} \mathscr{F}[\underline{\phi}_i] = \sum_{i=1}^{n} \mathscr{F}[\underline{\gamma}_i] = \sum_{i=1}^{n} \lambda_i$$

This is in fact true, but it does not follow so simply because in this procedure the maximization at each stage depends on the previous stages. The fact that it is true depends on the character of the surface, and it follows from Fan's Theorem (section 2.4).

3.4 MAXIMUM SEPARATION PROPERTY

There is another geometric property associated with the solution to the problem of section 3.2. Let Γ be an m-dimensional linear vector space the elements of which are functions over a certain time interval Ω. Suppose that the random process x(t) consists only of certain waveforms $s_1(t), \ldots, s_n(t)$ which occur with probabilities $P_1, \ldots, P_n$. Only one waveform occurs per trial. The autocorrelation function is, then, $R_x(s,t) = \sum_{i=1}^{n} P_i s_i(s)\, s_i(t)$, and we shall assume that $E[x(t)] = 0$.

Suppose that we arbitrarily pick a set of ℓ orthonormal functions $\gamma_1(t), \ldots, \gamma_\ell(t)$ which define an ℓ-dimensional hyperplane Γ_ℓ of Γ. Let $\gamma_{\ell+1}(t), \ldots, \gamma_m(t)$ be an arbitrary completion of the set so that $\{\gamma_i(t)\}$ is a basis for the whole space. The projections of the waveforms on Γ_ℓ are, then,

$$s_i'(t) = \sum_{j=1}^{\ell} \gamma_j(t) \int_\Omega s_i(t)\, \gamma_j(t)\, dt = \sum_{j=1}^{\ell} s_{ij} \gamma_j(t) \qquad i = 1, \ldots, n$$

where

$$s_{ij} = \int_\Omega s_i(t)\, \gamma_j(t)\, dt \qquad j = 1, \ldots, m$$

We shall define the average separation S of the $s_i'(t)$ in Γ_ℓ to be

$$S = \sum_{i,j=1}^{n} P_i P_j \int_\Omega [s_i'(t) - s_j'(t)]^2\, dt$$

and we shall be interested in finding what orientation of Γ_ℓ maximizes S. We have

$$S = \sum_{i,j=1}^{n} \sum_{k=1}^{\ell} P_i P_j (s_{ik} - s_{jk})^2$$

$$= \sum_{i,j=1}^{n} \sum_{k=1}^{\ell} P_i P_j s_{ik}^2 + \sum_{i,j=1}^{n} \sum_{k=1}^{\ell} P_i P_j s_{jk}^2 - 2 \sum_{i,j=1}^{n} \sum_{k=1}^{\ell} P_i P_j s_{ik} s_{ij}$$

$$= 2 \sum_{i=1}^{n} \sum_{k=1}^{\ell} P_i s_{ik}^2 - 2 \sum_{k=1}^{\ell} \left\{ \sum_{i=1}^{n} P_i s_{ik} \right\}^2$$

We note that

$$E[x(t)] = \sum_{i=1}^{n} P_i s_i(t) = \sum_{i=1}^{n} P_i \sum_{j=1}^{n} s_{ij} \gamma_j(t)$$

$$= \sum_{j=1}^{m} \gamma_j(t) \sum_{i=1}^{n} P_i s_{ij} = 0$$

therefore

$$\sum_{i=1}^{n} P_i s_{ij} = 0 \qquad j = 1, \ldots, m$$

and

$$S = 2 \sum_{i=1}^{n} \sum_{k=1}^{\ell} P_i s_{ik}^2$$

$$= 2 \sum_{i=1}^{n} \sum_{k=1}^{\ell} P_i \int_\Omega \int_\Omega s_i(s)\, s_i(t)\, \gamma_k(s)\, \gamma_k(t)\, ds\, dt$$

$$= 2 \sum_{k=1}^{\ell} \int_\Omega \int_\Omega R_x(s,t)\, \gamma_k(s)\, \gamma_k(t)\, ds\, dt$$

As we have seen before, this is maximized by using the first eigenfunctions of $R_x(s,t)$ for the $\gamma_1(t), \ldots, \gamma_\ell(t)$, so that the orientation of Γ_ℓ which maximizes the average separation is determined by these.

Consequently, we see that if we have a cluster of signals in function space, the orientation of the hyperplane which minimizes the error of representation in the lower dimension also maximizes the spread of the projection of this cluster on the hyperplane, weighted by the probabilities of occurrence of the signals. If there were some uncertainty as to the position of the signal points in the function space, then we might say that this orientation is the orientation of least confusion among the projections of the signal points on the hyperplane.

3.5 THE ASYMPTOTIC BEHAVIOR OF THE AVERAGE ERROR IN THE STATIONARY CASE

In this section we shall consider the representation of a stationary process x(t) for all time (see Jordan[16]). This will be done by dividing time into intervals of length 2A and using the optimum representation of section 3.2 for each interval. Since the process is stationary, the solution will be the same for each interval.

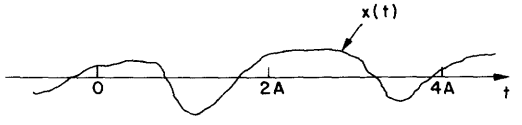


Fig. 5. Division of the process into intervals.

Suppose that we use n terms to represent each interval. We then define the density as $k = n/2A$, or the average number of terms per unit time. If we consider an interval of length 4A, as shown in Fig. 5, consisting of two subintervals of length 2A each separately represented, we would have an average error

$$\frac{2\theta^*(2A)}{4A} = \frac{\theta^*(2A)}{2A}$$

If we now increase the interval of representation to 4A while using 2n terms, that is, holding the density constant, we would have an average error $\frac{\theta^*(4A)}{4A}$. It is certainly true that

$$\frac{\theta^*(4A)}{4A} \leq \frac{\theta^*(2A)}{2A} \tag{16}$$

since if it were not true, this would contradict the fact that the representation is optimum. It is the object of this section to study the behavior of $\frac{\theta^*(2A)}{2A}$ as A increases while the density is held constant.

Since the process is stationary, $R_x(t,t) = R_x(0)$, and, from (14), we have

$$\frac{1}{2A}\theta^*(2A) = R_x(0) - \frac{1}{2A}\sum_{i=1}^{n}\lambda_i$$

where the λ_i are the eigenvalues of

$$\int_{-A}^{A} R_x(s-t)\,\phi_i(t)\,dt = \lambda_i \phi_i(s) \qquad -A \leq s \leq A$$

Since $n = 2kA$ must be a positive integer, A can only take on the values

$$A_n = \frac{n}{2k} \qquad n = 1, 2, \ldots$$

The sequence $\dfrac{\theta^*(2A_n)}{2A_n}$ is monotonically decreasing because of the argument leading to (16). Since $\theta^*(2A_n) \geq 0$, all n, the sequence must have a limit.[50] We want to find

$$\lim_{n\to\infty} \frac{\theta^*(2A_n)}{2A_n} = R_x(0) - \lim_{n\to\infty} \frac{1}{2A_n} \sum_{i=1}^{2kA_n} \lambda_i$$

We now make use of a theorem that will be proved in Appendix B.

THEOREM II.

$$\lim_{n\to\infty} \frac{1}{2A_n} \sum_{i=1}^{2kA_n} \lambda_i = \lim_{n\to\infty} \frac{k}{n} \sum_{i=1}^{n} \lambda_i = \int_E S_x(f)\,df$$

where

$$S_x(f) = \int_{-\infty}^{\infty} R_x(t)\, e^{-j2\pi ft}\,dt$$

and is the power density spectrum of the process, and

$$E = [f;\, S_x(f) \geq \ell]$$

where ℓ is adjusted in such a way that

$$\mu[E] = k \tag{17}$$

(The notation $[f;\, S_x(f) \geq \ell]$ means "the set of all f such that $S_x(f) \geq \ell$." $\mu[E]$ denotes the measure of the set E, or length for our purposes.) Now since

$$R_x(t) = \int_{-\infty}^{\infty} S_x(f)\, e^{j2\pi ft}\,df$$

$$R_x(0) = \int_{-\infty}^{\infty} S_x(f)\,df$$

and

$$\lim_{n\to\infty} \frac{\theta^*(2A_n)}{2A_n} = \int_{-\infty}^{\infty} S_x(f)\, df - \int_E S_x(f)\, df = \int_{E'} S_x(f)\, df \tag{18}$$

where $E' = [f; S_x(f) < \ell]$.

In other words, we take the power density spectrum (see Fig. 6) and adjust ℓ in such a way that the length along f for which $S_x(f) \geq \ell$ is k and then integrate over all the remaining regions. This gives a lower bound for the average error and the bound is approached asymptotically as A increases.

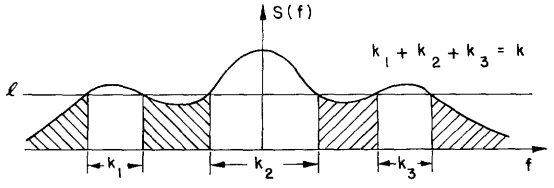


Fig. 6. Method of finding the asymptotic error.

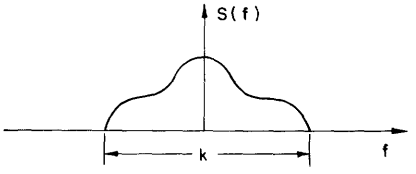


Fig. 7. Spectrum of a band-limited process.

If the process x(t) is bandlimited with bandwidth k/2 cps, that is, it has no power in the frequencies above k/2 cps, then we have the spectrum shown in Fig. 7. If we then use a density of k terms/sec, we see that ℓ must be adjusted, according to the condition of Eq. 16, to a level $\ell = 0$. By Eq. 17, we have

$$\lim_{n\to\infty} \frac{\theta^*(2A_n)}{2A_n} = \int_{E'} S_x(f)\, df = 0$$

This implies that we can approach arbitrarily closely an average error of zero with a finite time linear representation by allowing the time interval to become large enough. This is in agreement with the Sampling Theorem[28, 1] which states that x(t) can be represented exactly by R equally spaced samples per unit time; and, in addition, we are assured that this is the most efficient linear representation.

3.6 A RELATED CASE

Suppose that x(t) is a zero-mean random process in the interval [-A, A] with autocorrelation function $R_x(s,t)$. We now consider the problem in which the a_i are specified to be certain linear operations on x(t)

$$a_i = \int_{-A}^{A} x(t)\, g_i(t)\, dt \qquad i = 1, \ldots, n$$

and we minimize θ with F constrained as in section 3.2; that is,

$$F(t, a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \phi_i(t)$$

(c(t) = 0, since the process is zero mean). If we follow a straightforward minimization procedure, we find that the set $\{\phi_i(t)\}$ must satisfy

$$\int_{-A}^{A} R_x(t, s)\, g_i(s)\, ds = \sum_{j=1}^{n} \phi_j(t) \int\int_{-A}^{A} R_x(u, v)\, g_i(u)\, g_j(v)\, du\, dv \qquad i = 1, \ldots, n$$

which is just a set of linear equations in a parameter t.

If the a_i are samples of x(t), we then have $g_i(t) = \delta(t-t_i)$ and the set $\{\phi_i(t)\}$ is then the solution of

$$R_x(t, t_i) = \sum_{j=1}^{n} \phi_j(t)\, R_x(t_i, t_j) \qquad i = 1, \ldots, n \tag{19}$$

Solving this with matrix notation used, we have

$$\phi_j(t)] = [R_x(t_i, t_j)]^{-1}\, R_x(t, t_j)]$$

If we consider $\phi_j(t)]$ for $t = t_i$ (i=1, . . . , n), then we have the matrix equation

$$[\phi_j(t_i)] = [R_x(t_i, t_j)]^{-1}[R_x(t_i, t_j)] = [I]$$

where [I] denotes the identity matrix, so we see that

$$\phi_j(t) = \begin{cases} 1, & t = t_j \\ 0, & t = t_i,\ i \neq j \end{cases} \qquad j = 1, \ldots, n$$

If the process x(t) is stationary and the a_i are equally spaced samples in the interval $(-\infty, \infty)$, Eq. (19) becomes

$$R_x(t-kT_o) = \sum_{\ell=-\infty}^{\infty} \phi_\ell(t)\, R_x(kT_o - \ell T_o) \qquad k = 0, 1, -1, 2, -2 \ldots$$

where T_o is the period of sampling. Substituting $t' = t - kT_o$, we get

$$R_x(t') = \sum_{\ell=-\infty}^{\infty} \phi_\ell(t'+kT_o)\, R_x(kT_o - \ell T_o) \qquad k = 0, 1, -1, 2, -2 \ldots .$$

This holds for k equal to any integer, so that

$$R_x(t') = \sum_{\ell=-\infty}^{\infty} \phi_\ell(t'+(k+j)T_o)\, R_x((k+j)T_o - \ell T_o)$$

$$= \sum_{\ell=-\infty}^{\infty} \phi_{\ell+j}(t'+(k+j)T_o)\, R_x(kT_o - \ell T_o)$$

and we have

$$\phi_\ell(t+kT_o) = \phi_{\ell+j}(t+(k+j)T_o)$$

or

$$\phi_{\ell+j}(t+kT_o) = \phi_\ell(t+(k-j)T_o)$$

so that for $\ell = 0$, $k = 0$

$$\phi_j(t) = \phi_o(t-jT_o)$$

where $j = 0, 1, -1, 2, -2, \ldots$. The set $\{\phi_j(t)\}$ is just a set of translations of a basic interpolatory function, which is the solution of

$$R_x(t) = \sum_{\ell=-\infty}^{\infty} \phi_o(t'+\ell T_o)\, R_x(\ell T_o)$$

This problem has been studied by Tufts.[30] He has shown that the average error in this case is

$$R_x(0) - \int_{-\infty}^{\infty} \frac{[S_x(f)]^2}{\sum_{\ell=-\infty}^{\infty} S_x(f-\ell f_o)}\, df \tag{20}$$

where $f_o = 1/T_o$.

3.7 AN EXAMPLE

Let x(t) be a stationary random process in the interval $[-A, A]$ with a zero mean and autocorrelation function

$$R_x(s,t) = R_x(s-t) = \pi\, e^{-2\pi|s-t|}$$

The power density spectrum is then

$$S_x(f) = \frac{1}{1+f^2}$$

The eigenfunctions for this example are

$$\phi_i(t) = \begin{cases} c_i \cos b_i t & i, \text{ odd} \\ c_i \sin b_i t & i, \text{ even} \end{cases}$$

where the c_i are normalizing constants and the b_i are the solutions of the transcendental equations

$$b_i \tan b_i A = 2\pi \qquad i, \text{ odd}$$

$$b_i \cot b_i A = -2\pi \qquad i, \text{ even}$$

The eigenvalues are given by

$$\lambda_i = \frac{4\pi^2}{b_i^2 + 4\pi^2}$$

The details of the solution of this problem have been omitted because they may be found elsewhere.[52]

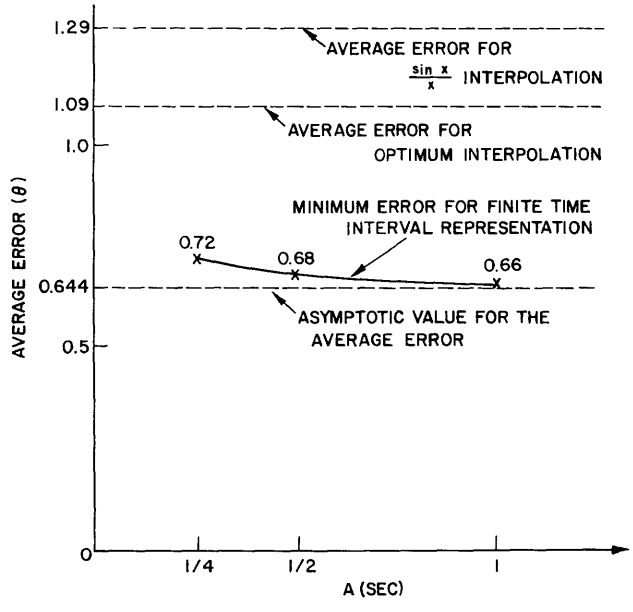


Fig. 8. Comparison of errors for several representations.

The minimum average error, $\frac{1}{2A}\theta^*$, has been computed for several values of A for the case k = 6 terms/sec, and these results are shown in Fig. 8. The predicted asymptotic value is

$$2\int_{k/2}^{\infty} S_x(f)\,df = 2\int_{3}^{\infty} \frac{1}{1+f^2}\,df = 0.644 \qquad (21)$$

This is plotted in Fig. 8 along with the error incurred by sampling at the rate of 6 samples/sec and reconstructing with $\frac{\sin x}{x}$ interpolatory functions. This error is just twice the error given in Eq. 31, or twice the area under the tails of the spectrum for $|f| > 3$. This is a known fact; but a short proof will be given in Appendix C for reference. Also shown in Fig. 8 is the error acquired by sampling and using an optimum interpolatory function. This error was computed from Eq. (20).

3.8 OPTIMIZATION WITH UNCERTAINTY IN THE REPRESENTATION

It is of interest to know whether or not the solution of section 3.2 is still optimum when the representation in the form of the set of random variables $\{a_i\}$ is subject to uncertainties. This would occur, for example, if the representation is transmitted over a noisy channel in some communication system.

In the following discussion we shall assume that the process is zero-mean; the representation is derived by the linear operations

$$a_i = \int_\Omega x(t)\, g_i(t)\, dt, \tag{22}$$

and the approximation function is

$$F(t, a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \phi_i(t).$$

Our object is, then, to determine under what conditions

$$\theta = E\left[\int_\Omega \left[x(t) - \sum_{i=1}^{n} (a_i+\epsilon_i)\, \phi_i(t)\right]^2 dt\right]$$

is minimized, when the ϵ_i are random variables representing the uncertainties. Under the assumption that $\{\phi_i(t)\}$ is an orthonormal set, we obtain

$$\theta = \int_\Omega R_x(t,t)\, dt - 2E\left[\sum_{i=1}^{n} (a_i+\epsilon_i) \int_\Omega x(t)\, \phi_i(t)\, dt\right] + \sum_{i=1}^{n} E(a_i+\epsilon_i)^2$$

and we substitute Eq. 22 to obtain

$$\theta = \int_\Omega R_x(t,t)\, dt - 2\sum_{i=1}^{n} \int_\Omega\int_\Omega R_x(s,t)\, g_i(s)\, \phi_i(t)\, ds\, dt + \sum_{i=1}^{n} \int_\Omega\int_\Omega R_x(s,t)\, g_i(s)\, g_i(t)\, ds\, dt$$

$$- 2\sum_{i=1}^{n} \int_\Omega E[\epsilon_i x(t)]\, \phi_i(t)\, dt + 2\sum_{i=1}^{n} \int_\Omega E[\epsilon_i x(t)]\, g_i(t)\, dt + \sum_{i=1}^{n} E[\epsilon_i^2]. \tag{23}$$

If we replace g_i by $g_i + \alpha\eta_i$ in this expression, we know from the calculus of variations[53] that a necessary condition that θ be a minimum with respect to g_i is

$$\frac{\partial}{\partial \alpha}\theta\Big|_{\alpha=0} \equiv 0.$$

Applying this value, we obtain

$$\frac{\partial}{\partial \alpha}\theta\Big|_{\alpha=0} = 2\int_\Omega \eta_i(s)\,ds\left\{\int_\Omega R_x(s,t)\,\phi_i(t)\,dt - \int_\Omega R_x(s,t)\,g_i(t)\,dt - E[\epsilon_i x(s)]\right\} = 0$$

and since $\eta_i(s)$ is arbitrary, the condition becomes

$$\int_\Omega R_x(s,t)[\phi_i(t)-g_i(t)]\,dt = E[\epsilon_i x(s)] \qquad \begin{matrix} s \in \Omega \\ i = 1,\ldots,n \end{matrix}$$

It is seen, then, that if $E[\epsilon_i x(s)] = 0$, $s \in \Omega$, then $\phi_i(t) = g_i(t)$ $(i=1,\ldots,n)$ satisfies the condition. (If $R_x(s,t)$ is positive definite, this solution is unique.) For this case, Eq. (23) becomes

$$\theta = \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n}\int_\Omega\int_\Omega R_x(s,t)\,\phi_i(s)\,\phi_i(t)\,ds\,dt + \sum_{i=1}^{n} E[\epsilon_i^2].$$

Consequently, we see that if $E[\epsilon_i x(s)] = 0$ $(i=1,\ldots,n)$, for all $s \in \Omega$, then the solution of section 3.2 is still optimum and the minimum error now is

$$\theta^* = \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n} E[\epsilon_i^2].$$

3.9 A MORE GENERAL NORM

Although in the general formulation of the problem given in section 3.1 we consider a general norm, up to this point we have made use of only the rms norm. In many problems, however, we shall be interested in a measure not only of the average difference between functions but also of other characteristics of the functions. For example, in Section I we described in what way a linear representation of the past of a random process is useful in a characterization of nonlinear systems. For the most part, such a characterization is useful only for those nonlinear systems for which the influence of the remote past on the operation of the system is small compared with the influence of the immediate past. In such a case, we would be interested not in a norm that weights the average difference between functions uniformly over function space, as in the rms norm, but in a norm that weights the immediate past more heavily than the remote past.

In this problem we might also be interested in a norm that discriminates not only in time but also in frequency. The high frequencies may influence the operation of the system to a lesser degree than the low frequencies. So, we see that it would be of interest to consider a norm that is more general than the rms norm which discriminates neither in time nor in frequency.

We consider here a generalization on the rms norm which allows more general discrimination in the characteristics of the random function. This norm has the additional property that with it the solution of the representation problem still requires only the solution of linear equations. This norm is

$$||f(t)|| = \left[\int_\Omega f_1^2(t)\ dt\right]^{1/2}$$

Here, $f_1(t)$ is obtained by operating linearly on f(t); that is,

$$f_1(t) = \int_\Omega K(t, u)\ f(u)\ du; \qquad t \in \Omega \tag{24}$$

where K(t, u) is determined by the requirements of the problem. (We have assumed that the linear operation is an integral operation, although this is not necessary. In our first special case below it will not be strictly an integral operation.) Essentially, what we have done is to pass the error e(t) through a linear filter and then use the root mean square. In order for this to be a true norm, K(t, u) must satisfy the condition

$$\int_\Omega K(t, u)\ f(u)\ du = 0 \qquad t \in \Omega \tag{25}$$

if and only if f(u) = 0 for $u \in \Omega$. (See sec. 2.1.) A necessary and sufficient condition that this be true is that the symmetrical kernel

$$K_1(s, t) = \int_\Omega K(u, s)\ K(u, t)\ du$$

be positive definite. This is because the conditions

$$f_1(t) = \int_\Omega K(t, u)\ f(u)\ du = 0$$

and

$$\int_\Omega f_1^2(t)\ dt = \int_\Omega \int_\Omega \left\{\int_\Omega K(t, u)\ K(t, v)\ dt\right\} f(u)\ f(v)\ du\ dv = 0$$

are equivalent.

The error, θ, now becomes

$$\theta = E\left[\int_\Omega dt \left\{\int_\Omega K(t,u)\left[x(u) - c(u) - \sum_{i=1}^{n} a_i\phi_i(u)\right] du\right\}^2\right]$$

$$= E\left[\int_\Omega dt \left\{\int_\Omega K(t,u)\, x(u)\, du - \int_\Omega K(t,u)\, c(u)\, du - \sum_{i=1}^{n} a_i \int_\Omega K(t,u)\, \phi_i(u)\, du\right\}^2\right]$$

so we see from the second of these equations that the problem reduces to the representation of the process

$$y(t) = \int_\Omega K(t,u)\, x(u)\, du$$

by the method of section 3.2. Consequently, our solution is

$$F^*(t, a_1, \ldots, a_n) = m_x(t) + \sum_{i=1}^{n} a_i\gamma_i(t) \tag{26}$$

in which the $\gamma_i(t)$ are solutions of

$$\Phi_i(t) = \int_\Omega K(t,u)\, \gamma_i(u)\, du \tag{27}$$

and the $\Phi_i(t)$ are the eigenfunctions of

$$\int_\Omega G(s,t)\, \Phi_i(t)\, dt = \lambda_i\Phi_i(s) \qquad s \in \Omega \tag{28}$$

arranged in the order $\lambda_1 \geq \lambda_2 \geq \ldots$. $G(s,t)$ is found from

$$G(s,t) = \int_\Omega\int_\Omega K(s,u)\, K(t,v)\, r_x(u,v)\, du\, dv \tag{29}$$

and we have

$$a_i^* = \int_\Omega ds\Phi_i(s) \int_\Omega K(s,v)[x(v)-m_x(v)]\, dv.$$

The minimum error is

$$\theta^* = \int_\Omega G(t,t)\, dt - \sum_{i=1}^{n} \int_\Omega\int_\Omega G(s,t)\, \Phi_i(s)\, \Phi_i(t)\, ds\, dt$$

$$= \int G(t,t)\, dt - \sum_{i=1}^{n} \lambda_i \tag{30}$$

in which the λ_i are the eigenvalues of Eq. (28).

We have a particularly simple case when $K(s,t)$ is expressed over the basis of eigenfunctions $\{\psi_i(s)\}$ of $r_x(s,t)$; that is,

$$K(s,t) = \sum_{i=1}^{\infty} \beta_i \psi_i(s)\ \psi_i(t).$$

We then have for $G(s,t)$

$$G(s,t) = \sum_{i=1} \beta_i^2 \alpha_i \psi_i(s)\ \psi_i(t)$$

where the α_i are the eigenvalues of $r_x(s,t)$. We then have

$$\Phi_i(s) = \psi_i(s)$$

$$\lambda_i = \beta_i^2 \alpha_i$$

$$\gamma_i(s) = \frac{1}{\beta_i} \psi_i(s)$$

for $i = 1, 2, \ldots$.

We shall now discuss two special cases of this norm which demand special attention.

THE FIRST CASE

First, we consider the rms norm weighted in time, that is, we have

$$\|f(t)\| = \left[\int_\Omega W^2(t)\ f^2(t)\ dt\right]^{1/2}$$

so that the linear operation is just multiplication by $W(t)$. This corresponds to a kernel $K(t,u) = W(t)\ \delta(t-u)$. The solution now is

$$F^*(t, a_1, \ldots, a_n) = m_x(t) + \sum_{i=1}^{n} a_i \frac{\Phi_i(t)}{W(t)}$$

$$\int_\Omega W(s)\ r_x(s,t)\ W(t)\ \Phi_i(t)\ dt = \lambda_i\ \Phi_i(s) \qquad s \in \Omega$$

$$a_i^* = \int_\Omega W(t)\ \Phi_i(t)[x(t)-m_x(t)]\ dt$$

in which the error is

$$\theta^* = \int_\Omega W^2(t)\, r_x(t,t)\, dt - \sum_{i=1}^{n} \int_\Omega \int_\Omega W(s)\, r_x(s,t)\, W(t)\, \Phi_i(s)\, \Phi_i(t)\, ds\, dt$$

$$= \int_\Omega W^2(t)\, r_x(t,t)\, dt - \sum_{i=1}^{n} \lambda_i .$$

This is of special interest in the nonlinear filter problem in which we want to represent the past of a random function with a norm that minimizes the effect of the remote past. In fact, if the process is stationary, we must use this weighted norm in order to get an answer to the problem at all. This is because if we use the method of section 3.2, the first term of Eq. (14) would be infinite; that is,

$$\int_{-\infty}^{0} r_x(0)\, dt = \infty$$

and no matter how many terms we use, we would not improve the situation. Also, the kernel of the integral equation

$$\int_{-\infty}^{0} r_x(s-t)\, \gamma_i(t)\, dt = \lambda_i \gamma_i(s) \qquad s \in [-\infty, 0]$$

is not of integrable square; that is, we have

$$\int\int_{-\infty}^{0} |r_x(s-t)|^2\, ds\, dt = \infty$$

so that we are not assured that the integral equation has a countable set of solutions. However, if we use a weighting function W(t) chosen in such a way that

$$\int_{-\infty}^{0} W^4(t)\, r_x^2(0)\, dt = r_x^2(0) \int_{-\infty}^{0} W^4(t)\, dt$$

is finite, then we can find a solution.

It might be well to point out also that although we have said that we must pick a weighting W(t), we have not attempted to suggest what W(t) to use. This must depend upon the nonlinear filtering problem at hand and upon the insight and judgment of the designer.

As an example we consider the zero-mean random process x(t) with autocorrelation function $R_x(s,t) = e^{-|s-t|}$. We shall be interested in representing the past of this process with a weighting function $W(t) = e^{t}$ over $[-\infty, 0]$. However, for the sake of convenience, we shall use the interval $[0, \infty]$ and weighting function $W(t) = e^{-t}$. In this case the solutions of the integral equation[17]

$$\int_0^\infty e^{-s}\, e^{-|s-t|}\, e^{-t}\, \Phi_i(t)\, dt = \lambda_i \Phi_i(s) \qquad s \geq 0$$

are

$$\Phi_i(t) = A_i\, e^{-t}\, J_1\left[\sqrt{\frac{2}{\lambda_i}}\, e^{-t}\right]$$

$$\lambda_i = \frac{2}{q_i^2} \tag{31}$$

Here, the q_i are the positive roots of $J_o(q_i) = 0$. The $J_i(x)$ are the Bessel functions of the first order, and the A_i are normalizing constants. The error in this case is

$$\theta^* = \int_0^\infty e^{-2t}\, dt - \sum_{i=1}^{n} \lambda_i = \frac{1}{2} - \sum_{i=1}^{n} \lambda_i$$

The first two zeros of $J_o(x)$ are[54]

$$q_1 = 2.4048$$

$$q_2 = 5.5201$$

so that the first two eigenvalues are

$$\lambda_1 = 0.3458$$

$$\lambda_2 = 0.0656$$

The error for one term is then

$$\theta_1^* = 0.5 - 0.3458 = 0.1542 \tag{32}$$

and for two terms

$$\theta_2^* = 0.5 - 0.3458 - 0.0656 = 0.0886 \tag{33}$$

THE SECOND CASE

Second, consider the case in which the interval of interest is $[-\infty, \infty]$, and the kernel of the linear operation of Eq. (24) factors into the form

$$K(s,u) = K_1(s)\, K_2(s-u)$$

so that we have

$$f_1(s) = K_1(s) \int_{-\infty}^{\infty} K_2(s-u)\, f(u)\, du$$

Thus, the operation consists of a cascade of stationary linear filtering and multiplication. If $K_1(s) \geq 0$ and the Fourier transform of $K_2(s)$ is real and positive, then we can consider the norm as a frequency weighting followed by a time weighting. (For these conditions, the condition of Eq. (25) for the kernel of the norm is also satisfied.)

Let us consider the example of the representation of white noise x(t) of autocorrelation function $R_x(s, t) = \delta(s-t)$ and mean zero. Here we use as weightings

$$K_1(s) = e^{-s^2}$$

$$K_2(s) = e^{-s^2}$$

that is, Gaussian weightings both in time and frequency. From Eqs. (28) and (29) we see that we must find the eigenfunctions of G(s, t), where

$$G(s, t) = \int\int_{-\infty}^{\infty} K_1(s)\, K_2(s-u)\, K_1(t)\, K_2(t-v)\, R_x(u, v)\, du\, dv$$

$$= \int\int_{-\infty}^{\infty} e^{-s^2}\, e^{-(s-u)^2}\, e^{-t^2}\, e^{-(t-v)^2}\, \delta(u-v)\, du\, dv$$

$$= e^{-s^2}\, e^{-t^2} \int_{-\infty}^{\infty} e^{-(s-u)^2}\, e^{-(t-u)^2}\, du$$

The Fourier transform of e^{-t^2} (see Appendix D) is

$$\int_{-\infty}^{\infty} e^{-t^2}\, e^{-j2\pi ft}\, dt = \pi\, e^{-\pi^2 f^2}$$

we know that

$$\int_{-\infty}^{\infty} f(\sigma)\, g(t-\sigma)\, d\sigma = \int_{-\infty}^{\infty} F(f)\, G(f)\, e^{j2\pi ft}\, df$$

where F(f) and G(f) are the Fourier transforms of f(t) and g(t). We then have

$$\int_{-\infty}^{\infty} e^{-u^2}\, e^{-(s-t-u)^2}\, du = \pi \int_{-\infty}^{\infty} e^{-2\pi^2 f^2}\, e^{j2\pi f(s-t)}\, dt$$

$$= \sqrt{\frac{\pi}{2}}\, e^{-\frac{1}{2}(s-t)^2}$$

so that

$$G(s, t) = \sqrt{\frac{\pi}{2}}\, e^{-s^2}\, e^{-t^2}\, e^{-\frac{1}{2}(s-t)^2}.$$

It is shown in Appendix D that the eigenfunctions of this kernel are

$$\Phi_i(t) = A_i\, e^{\sqrt{2}t^2}\, \frac{d^i}{dt^i}\, e^{-2\sqrt{2}t^2} \qquad i = 0, 1, \ldots$$

Here, the A_i are normalizing constants and the eigenvalues are

$$\lambda_i = \pi \sqrt{\frac{1}{3 + 2\sqrt{2}}}\, (3-2\sqrt{2})^i \qquad i = 0, 1, \ldots$$

It is seen that these functions are the Hermite functions modified by a scale factor. The Hermite functions[55] are given by

$$H_n(t) = (2^n n!\, \sqrt{\pi})^{-1/2}\, e^{t^2/2}\, \frac{d^n}{dt^n}\, e^{-t^2} \qquad n = 0, 1, 2, \ldots$$

Therefore, we have

$$\Phi_i(t) = (2\sqrt{2})^{1/4}\, H_i[(2\sqrt{2})^{1/2} t] \qquad i = 0, 1, 2, \ldots$$

and the A_i are given by

$$A_i = \frac{(2\sqrt{2})^{1/4}}{(2^i i!\, \sqrt{\pi})^{1/2}} \qquad i = 0, 1, 2, \ldots$$

Referring to Eq. (26) we see that in order to have the complete solution we must find the $\gamma_i(t)$ that are the solutions of

$$\Phi_i(t) = \int_{-\infty}^{\infty} e^{-t^2}\, e^{-(t-u)^2}\, \gamma_i(u)\, du \qquad i = 0, 1, 2, \ldots$$

according to Eq. (27). It is shown in Appendix E that the solution is

$$\gamma_i(t) = A_i \frac{(\sqrt{2}+2)^i}{\sqrt{\pi(2-\sqrt{2})}}\, e^{\frac{1+\sqrt{2}}{2+\sqrt{2}} t^2}\, \frac{d^i}{dt^i}\, e^{-\sqrt{2}t^2}$$

so that the best representation is given by

$$F^*(t, a_o, \ldots, a_n) = \sum_{i=0}^{n} a_i A_i \frac{(\sqrt{2}+2)^i}{\sqrt{\pi(2-\sqrt{2})}}\, e^{\frac{1+\sqrt{2}}{2+\sqrt{2}} t^2}\, \frac{d^i}{dt^i}\, e^{-\sqrt{2}t^2}$$

and

$$a_i^* = \int_{-\infty}^{\infty} ds\; A_i\, e^{(\sqrt{2}-1)s^2}\, \frac{d^i}{ds^i}\, e^{-2\sqrt{2}s^2} \int_{-\infty}^{\infty} e^{-(s-t)^2}\, x(t)\, dt$$

and the error is

$$\theta^* = \int_{-\infty}^{\infty} \sqrt{\frac{\pi}{2}}\, e^{-2t^2}\, dt - \sum_{i=0}^{n} \pi \sqrt{\frac{1}{3 + 2\sqrt{2}}}\, (3-2\sqrt{2})^i$$

$$= \frac{\pi}{2} - \sum_{i=0}^{n} \frac{\pi}{2} \sqrt{\frac{4}{3 + 2\sqrt{2}}}\, (3-2\sqrt{2})^i$$

3.10 COMPARISON WITH LAGUERRE FUNCTIONS

We now return to the first example of section 3.9, but this time we use Laguerre functions in place of the functions of Eq. (31). We shall be interested in just how close we can come to the minimum possible error given by Eqs. (32) and (33). The Laguerre functions[56] are given by

$$L_{n+1}(x) = \frac{1}{n!}\, e^{x/2}\, \frac{d^n}{dx^n}\, (x^n e^{-x}) \qquad n = 0, 1, 2, \ldots \tag{34}$$

for $x \geq 0$.

Since orthogonality of functions over $[0, \infty]$ is invariant with respect to a change in scale, we have a degree of freedom at our disposal. The Laguerre functions given above satisfy the relation

$$\int_0^{\infty} L_i(y)\, L_j(y)\, dy = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and if we make the change of variable $y = \alpha x$, we have

$$\alpha \int_0^{\infty} L_i(\alpha x)\, L_j(\alpha x)\, dx = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

from which it follows that the set of functions $\sqrt{\alpha}\, L_n(\alpha x)$ is orthonormal. We shall be interested in picking a best scale factor α for a representation in terms of these functions.

By replacing the functions $\Phi_i(t)$ in Eq. (30) by the set $\sqrt{\alpha}\, L_n(\alpha x)$, we obtain for the error

$$\theta = \int_0^{\infty} W^2(t)\, r_x(t, t)\, dt - \sum_{i=1}^{n} \int_0^{\infty} \int_0^{\infty} W(s)\, r_x(s, t)\, W(t)\, \sqrt{\alpha}\, L_i(\alpha s)\, \sqrt{\alpha}\, L_i(\alpha t)\, ds\, dt$$

and for the example it becomes

$$\theta = \frac{1}{2} - \sum_{i=1}^{n} \int_0^\infty \int_0^\infty e^{-s-t} e^{-|s-t|} \, aL_i(as) \, L_i(at) \, ds \, dt.$$

Suppose that n = 1. The first scaled Laguerre function is

$$\sqrt{a} \, L_1(ax) = \sqrt{a} \, e^{-\frac{a}{2}x}$$

so that we have for the error

$$\theta_1(a) = \frac{1}{2} - \int_0^\infty \int_0^\infty e^{-s-t-|s-t|} \, a \, e^{-\frac{as}{2} - \frac{at}{2}} \, ds \, dt$$

which on performing the integration becomes

$$\theta_1(a) = \frac{1}{2} - \frac{4a}{(a+2)(a+4)}.$$

This error is shown in Fig. 9 plotted as a function of a. It has a minimum at $a = 2\sqrt{2}$ for which $\theta_1(2\sqrt{2}) = 0.157$.

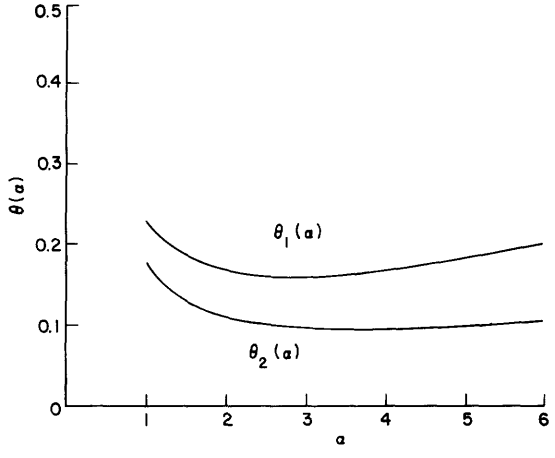


Fig. 9. The error as a function of scale factor for Laguerre functions.

Now suppose that n = 2. The second scaled Laguerre function is

$$\sqrt{a} \, L_2(ax) = \sqrt{a}(e^{-ax/2} - ax \, e^{-ax/2})$$

and the error becomes

$$\theta_2(a) = \frac{1}{2} - \frac{4a}{(a+2)(a+4)}$$

$$- \int_0^\infty \int_0^\infty e^{-s-t-|s-t|} \, a[e^{-as/2} - as \, e^{-as/2}][e^{-at/2} - at \, e^{-at/2}] \, ds \, dt$$

which on performing the integration becomes

$$\theta_2(a) = \frac{1}{2} - \frac{4a}{(a+2)(a+4)} - \frac{4a(a^3-4a-16)}{(a+2)^3(a+4)^2}$$

$$= \frac{1}{2} - \frac{-a^5 + 16a^3 - 32a + 128}{(a+2)^4(a+4)^3} .$$

This is also shown in Fig. 9 and it is minimum at $a = 4$, for which $\theta_2(4) = 0.093$.

We see, first of all, that the best scale factor for n = 1 is not the same as it is for n = 2. Also, it is interesting that the performance of the Laguerre functions for the best scale factor is remarkably close to optimum. The minima of the curves in Fig. 9 are very nearly the values given in Eqs. (32) and (33).

This example illustrates the value of knowing the optimum solution. In practice, if we are interested in representing the past of x(t), we would derive the random variables a_i from x(t) by means of linear filters. In this example, the synthesis of the filters for the optimum solution would be much more difficult than the synthesis of the filters for Laguerre functions. For representing the past of x(t) we would have (reversing the sign of t since in the example we have used the interval $[0, \infty]$)

$$a_i^* = \int_{-\infty}^{0} x(t)\, W(-t)\, \Phi_i(-t)\, dt$$

$$= \int_{-\infty}^{0} x(t)\, e^t\, A_i\, e^t\, J_1\left[\sqrt{\frac{2}{\lambda_i}}\, e^t\right] dt$$

so that we would use a linear filter of impulse response

$$h_i(t) = A_i\, e^{-2t}\, J_1\left[\sqrt{\frac{2}{\lambda_i}}\, e^{-t}\right]$$

which would not be easy to synthesize. Now, if we use Laguerre functions we would have

$$a_i = \int_{-\infty}^{0} x(t)\, e^t \sqrt{a}\, L_i(-at)\, dt$$

and we would use a filter of impulse response

$$h_i(t) = \sqrt{a}\, e^{-t}\, L_i(at) \tag{35}$$

which is quite easy to synthesize and gives us an error very close to optimum. By means of cascades of simple linear networks we can synthesize impulse responses in the form of Laguerre functions[23] or other orthonormal sets of the exponential type.[11] In Eq. (35) we have a multiplying factor of e^{-t} which can be accounted for in the complex

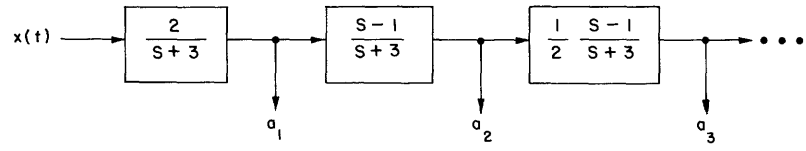


Fig. 10. A linear circuit for representing the past of a signal.

plane by displacing the poles and zeros of these networks in the direction of the real axis by -1. For example, suppose that we want to represent the past of x(t) using Laguerre functions with a scale factor $\alpha = 4$. By observing Eq. (34), we see that the Laplace transform of a Laguerre function is

$$\mathscr{L}_{n+1}(s) = \frac{1}{n!} \frac{\left(s-\frac{1}{2}\right)^n}{\left(s+\frac{1}{2}\right)^{n+1}} \qquad n = 0, 1, \ldots \tag{36}$$

so that the Laplace transform of $h_i(t)$, from Eq. (35), is

$$H_{n+1}(s) = \frac{2}{n!} \frac{(s-1)^n}{(s+3)^{n+1}} \qquad n = 0, 1, \ldots$$

We then see that we could derive the random variables a_i from the past of x(t) by using the cascade of linear networks shown in Fig. 10.

By replacing s by $j2\pi f$ in Eq. (36), we obtain the Fourier transform of $L_n(t)$ which is

$$\frac{1}{n!} \frac{\left(j2\pi f-\frac{1}{2}\right)^n}{\left(j2\pi f+\frac{1}{2}\right)^{n+1}} = \frac{1}{n!} \frac{1}{j2\pi f+\frac{1}{2}} \left[\frac{j2\pi f-\frac{1}{2}}{j2\pi f+\frac{1}{2}}\right]^n$$

The magnitude squared of this expression is

$$\left(\frac{1}{n!}\right)^2 \frac{4}{1+16\pi^2 f^2} \tag{37}$$

We note that this is similar in form to the spectrum of x(t) in the example. That is, since the correlation function of x(t) was $R_x(t) = \exp(|t|)$, the spectrum was the Fourier transform of this, or $S_x(f) = \frac{2}{1+4\pi^2 f^2}$. Heuristically speaking, this may be the reason why the set of Laguerre functions did so well. If the spectrum of x(t) were markedly different from the form of Eq. (37), then we might not expect the results to be as good.

IV. REPRESENTATION IN THE PRESENCE OF NOISE AND ITS BEARING ON OPTIMUM LINEAR SYSTEMS

4.1 REPRESENTATION IN THE PRESENCE OF NOISE

There are, perhaps, many situations in which a representation of a random signal is desired when the signal is not available directly, but only in a more or less contaminated form. Such a situation would occur, for example, when the representation is derived from the signal after it has been transmitted over a noisy channel. In this section we shall deal primarily with this problem and its close relationship to optimum, time-variant linear systems.

A discrete representation of $x(t)$ will be found, but the set of random variables $\{a_i\}$ will be derived from another process $y(t)$ that is statistically dependent on $x(t)$. The process $y(t)$ will, in most cases, be the perturbed version of $x(t)$. In a fashion similar to that of section 3.1 we have

$$a_i = T_i[y(t)] \qquad i = 1, \ldots, n$$

$$z(t) = F(t, a_1, \ldots, a_n)$$

and the problem can be stated generally as the minimum problem

$$\min_{\{T_i\}} \min_{F} E\left[\| x(t) - F(t, a_1, \ldots, a_n) \|^2 \right].$$

We shall now consider the linear case in which we find it necessary not only to restrict $F(t, a_1, \ldots, a_n)$ to be of the form $c(t) + \sum_{i=1}^{n} a_i \phi_i(t)$ but also to restrict the functionals to be linear in $x(t)$. The latter restriction does not follow from the former as it did in the case of direct representation. Also, we shall assume that the processes are zero-mean; otherwise, it is only necessary to subtract the mean, as we saw in section 3.2. Making use of the same norm as before, we shall minimize

$$\theta = E\left[\int_\Omega \left[x(t) - \sum_{i=1}^{n} a_i \phi_i(t) \right]^2 dt \right] \tag{38}$$

and without loss of generality we can assume that

$$\int_\Omega \phi_i(t)\, \phi_j(t)\, dt = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Since the functionals are linear, we shall assume that

$$a_i = \int_\Omega g_i(t)\, y(t)\, dt \qquad i = 1, \ldots, n$$

Substituting this in Eq. (38), we have

$$\theta = E\Big[\int_\Omega x^2(t)\, dt - 2 \sum_{i=1}^{n} \int_\Omega x(s)\, \phi_i(s)\, ds \int_\Omega y(t)\, g_i(t)\, dt$$

$$+ \sum_{i=1}^{n} \int_\Omega \int_\Omega y(s)\, y(t)\, g_i(s)\, g_i(t)\, ds\, dt\Big]$$

and, after interchanging the order of averaging and integration, we obtain

$$\theta = \int_\Omega R_x(t,t)\, dt - 2 \sum_{i=1}^{n} \int_\Omega \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, g_i(t)\, ds\, dt$$

$$+ \sum_{i=1}^{n} \int_\Omega \int_\Omega R_y(s,t)\, g_i(s)\, g_i(t)\, ds\, dt. \tag{39}$$

Our object is then to minimize this with respect to the sets of functions $\{g_i(t)\}$ and $\{\phi_i(t)\}$ under the assumption that $\{\phi_i(t)\}$ is an orthonormal set. First, we minimize with respect to $g_i(t)$. Then we replace $g_i(t)$ by $g_i(t) + \alpha\eta_i(t)$ and solve the equation

$$\left.\frac{\partial\theta}{\partial\alpha}\right|_{\alpha=0} = 0$$

from which we obtain

$$\int_\Omega R_y(s,t)\, g_i(s)\, ds = \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds = f_i(t) \qquad t \in \Omega\ . \tag{40}$$

By Picard's theorem (see sec. 2.2), there exists an integrable square solution to Eq. (40) if and only if the series

$$\sum_{j=1}^{\infty} \frac{1}{\beta_j^2} \left[\int_\Omega e_j(t)\, f_i(t)\, dt\right]^2 \tag{41}$$

converges, where the β_i and $e_i(t)$ are the eigenvalues and eigenfunctions of $R_y(s,t)$. This solution is

$$g_i(s) = \sum_{j=1}^{\infty} \frac{1}{\beta_j}\, e_j(s) \int_\Omega e_j(t)\, f_i(t)\, dt. \tag{42}$$

This solution can be verified by substitution back in Eq. (40). We shall assume hereafter that $g_i(s)$ is given by Eq. (42).

By substituting Eq. (40) in Eq. (39), we obtain

$$\theta = \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n} \int_\Omega \int_\Omega R_{xy}(s,t)\,\phi_i(s)\,g_i(t)\,ds\,dt \tag{43}$$

and by substituting Eq. (42) in Eq. (43), we have

$$\theta = \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n} \int_\Omega \int_\Omega R_{xy}(s,t)\,\phi_i(s)\,ds \sum_{j=1}^{\infty} \frac{1}{\beta_j}\,e_j(t) \int_\Omega e_j(u)\,f_i(u)\,du\,dt$$

$$= \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n} \sum_{j=1}^{\infty} \frac{1}{\beta_j} \int_\Omega \int_\Omega R_{xy}(s,t)\,\phi_i(s)\,e_j(t)\,ds\,dt \int_\Omega \int_\Omega R_{xy}(v,u)\,\phi_i(v)\,e_j(u)\,dv\,du$$

If we set $h_j(s) = \int_\Omega R_{xy}(s,t)\,e_j(t)\,dt$, after some rearrangement we obtain

$$\theta = \int_\Omega R_x(t,t)\,dt - \sum_{i=1}^{n} \int_\Omega \int_\Omega K(s,v)\,\phi_i(s)\,\phi_i(v)\,ds\,dv \tag{44}$$

where

$$K(s,v) = \sum_{j=1}^{\infty} \frac{1}{\beta_j} h_j(s)\,h_j(v) = \sum_{j=1}^{\infty} \frac{1}{\beta_j} \int_\Omega R_{xy}(s,t)\,e_j(t)\,dt \int_\Omega R_{xy}(v,u)\,e_j(u)\,du.$$

We know, then, from the corollary of Theorem I (see sec. 2.4) that θ is minimized by choosing as the set $\{\phi_i(t)\}$, the set of eigenfunctions $\{\gamma_i(t)\}$ of the kernel $K(s,t)$. Our solution is

$$F^*(t, a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \gamma_i(t).$$

Here, the $\gamma_i(t)$ are the solutions of

$$\int_\Omega K(s,t)\,\gamma_i(t)\,dt = \lambda_i \gamma_i(s) \qquad s \in \Omega$$

arranged in the order $\lambda_1 \geq \lambda_2 \geq \ldots$ and

$$K(s,t) = \sum_{j=1}^{\infty} \frac{1}{\beta_j} \left[\int_\Omega R_{xy}(s,u)\,e_j(u)\,du\right]\left[\int_\Omega R_{xy}(t,v)\,e_j(v)\,dv\right] \tag{45}$$

in which β_i and $e_i(t)$ are the eigenvalues and eigenfunctions of $R_y(s,t)$, and

$$a_i^* = \int_\Omega g_i(t)\, y(t)\, dt \qquad i = 1, 2, \ldots . \tag{46}$$

Here, the $g_i(t)$ are solutions of

$$\int_\Omega R_y(s,t)\, g_i(s)\, ds = \int_\Omega R_{xy}(s,t)\, \gamma_i(s)\, ds = f_i(t) \qquad i = 1, 2, \ldots \tag{47}$$

and the error is

$$\theta^* = \int_\Omega R_x(t,t)\, dt - \sum_{i=1}^{n} \lambda_i .$$

Thus, we have found the solutions to our problem of representation in the presence of noise by using a mean-square norm. The solution for the more general norm discussed in section 3.9 can also be found in precisely the same way as before.

In finding this result, we have assumed that Eq. (40) has solutions $g_i(t)$ that are integrable square. The result can be proved, however, under slightly more general conditions. This condition is that the

$$f_i(t) = \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds$$

each be expressible as a uniformly convergent series of the eigenfunctions $e_i(t)$. This condition includes our original assumption, since if $g_i(t)$ is integrable square, $f_i(t)$ can be expressed as a uniformly convergent series of eigenfunctions of $R_y(s,t)$ (see sec. 2.2). In order to show that it is more general, let us consider the case in which $g_i(s)$ is the impulse function $\delta(s-s_1)$. We have $f_i(t) = R_y(s_1, t)$, and from Mercer's theorem

$$f_i(t) = \sum_{i=1}^{\infty} \beta_i e_i(s_1)\, e_i(t)$$

and the series converges uniformly.

For a positive definite, nondegenerate kernel, the order of summation and integration in Eq. (45) cannot be interchanged without sacrifice of rigor because the series

$$\sum_{j=1}^{\infty} \frac{1}{\beta_j}\, e_j(u)\, e_j(v)$$

does not converge either uniformly or in the mean. As we pointed out in section 2.3, this series can represent the operation that is inverse to the operation $z(s) = \int_\Omega R_y(s,t)\, f(t)\, dt$. We shall formally denote this series by

$$R_y^{-1}(s,t) = \sum_{j=1}^{\infty} \frac{1}{\beta_j} e_j(s)\, e_j(t).$$

With this notation, Eq. (45) becomes

$$K(s,t) = \int_\Omega \int_\Omega R_{xy}(s,u)\, R_{xy}(t,v)\, R_y^{-1}(u,v)\, du\, dv \tag{48}$$

(If the kernels are degenerate, this is equivalent to the matrix multiplication $[K] = [R_{xy}][R_y^{-1}][R_{xy}]^T$), and for Eq. (48) we have

$$g_i(t) = \int_\Omega \int_\Omega R_y^{-1}(t,s)\, R_{xy}(u,s)\, \gamma_i(u)\, du\, ds \tag{49}$$

which are to be interpreted only in a symbolic sense.

4.2 ANOTHER INTERPRETATION

We have found the solution in a manner that is more or less straightforward but still not very enlightening. We now consider a slightly different approach that will give us a better idea of what the solution means. Considering Eq. (40), we see that it implies that there exists some sort of linear relation between $g_i(t)$ and $\phi_i(t)$. We could write

$$g_i(t) = \int_\Omega h(s,t)\, \phi_i(s)\, ds \qquad i = 1, \ldots, n.$$

If we substitute this in Eq. (40), we obtain

$$\int_\Omega R_y(s,t) \int_\Omega h(u,s)\, \phi_i(u)\, du\, ds = \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds.$$

Then we interchange the order of integration

$$\int_\Omega \phi_i(u)\, du \int_\Omega h(u,s)\, R_y(s,t)\, ds = \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds$$

and, since we assume that the set $\{\phi_i(t)\}$ is complete, we must have

$$\int_\Omega h(u,s)\, R_y(s,t)\, ds = R_{xy}(u,t) \qquad u, t \in \Omega$$

which is similar to the integral equation of Booton[57] for the optimum time-variant filter. If we invert this equation formally, we obtain

$$h(u,s) = \int_\Omega R_y^{-1}(t,s)\, R_{xy}(u,t)\, dt.$$

If we pass y(t) through this filter, the output is

$$z(t) = \int_{\Omega} h(t,u)\, y(u)\, du.$$

The autocorrelation function of this output is, then,

$$R_z(s,t) = E[z(s)z(t)] = \int_{\Omega}\int_{\Omega} h(s,v)\, h(t,u)\, R_y(u,v)\, du\, dv$$

$$= \int_{\Omega} h(s,v)\, R_{xy}(t,v)\, dv$$

$$= \int_{\Omega}\int_{\Omega} R_{xy}(s,u)\, R_{xy}(t,v)\, R_y^{-1}(u,v)\, du\, dv$$

which is identical to the kernel given by Eq. (48). The solution can then be described in the following way. We first pass y(t) through an optimum linear filter, and then we represent the output in the optimal manner described in section 3.2.

Special Cases

Case 1. We consider the case in which the signal y(t) is white; that is, $R_y(s,t) = \delta(s-t)$. Observing that the kernel that is inverse to an impulse is also an impulse, we have for K(s,t)

$$K(s,t) = \int_{\Omega} R_{xy}(s,u)\, R_{xy}(t,u)\, du$$

so that if λ_i and $\gamma_i(s)$ are the eigenvalues and eigenfunctions of K(s,t), we have

$$F^*(t, a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \gamma_i(t)$$

$$a_i^* = \int_{\Omega} x(t)\, g_i(t)\, dt \qquad i = 1, \ldots, n$$

in which, by Eq. (47),

$$g_i(t) = \int_{\Omega} R_{xy}(s,t)\, \gamma_i(s)\, ds$$

and the error is

$$\theta^* = \int_{\Omega} R_x(t,t)\, dt - \sum_{i=1}^{n} \lambda_i.$$

Case 2. Now suppose that the signal y(t) is the original signal x(t) plus independent white noise so that $R_{xy}(s,t) = R_x(s,t)$ and $R_y(s,t) = R_x(s,t) + N_o\delta(s-t)$. From Eq. (46),

we have

$$\int_\Omega R_x(s,t)\, e_i(t)\, dt + N_o e_i(s) = \beta_i e_i(s)$$

from which we get

$$e_i(t) = \gamma_i(t)$$

$$\beta_i = a_i + N_o$$

where a_i and $\gamma_i(t)$ are the eigenvalues and eigenfunctions of $R_x(s,t)$. $K(s,t)$, from Eq. (45), is then

$$K(s,t) = \sum_{j=1}^{\infty} \frac{a_i^2}{a_i + N_o} \gamma_j(s)\, \gamma_j(t).$$

From Eq. (47) we have

$$\int_\Omega R_x(s,t)\, g_i(s)\, ds + N_o g_i(t) = a_i \gamma_i(t)$$

so that

$$g_i(t) = \frac{a_i}{a_i + N_o} \gamma_i(t)$$

and the results are

$$F^*(t, a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \gamma_i(t)$$

$$a_i^* = \frac{a_i}{a_i + N_o} \int_\Omega x(t)\, \gamma_i(t)\, dt \qquad i = 1, \ldots, n$$

$$\theta^* = \int_\Omega R_x(t,t)\, dt - \sum_{i=1}^{n} \frac{a_i^2}{a_i + N_o}.$$

4.3 A BODE-SHANNON APPROACH

The derivation of the main results of sections 4.1 and 4.2 were rather long-winded; however, we note that the results of the first special case are quite simple. We shall now describe how this result can be derived in a shorter, more heuristic way, and then we shall argue that any problem can be reduced to this one by passing the process y(t) through a whitening filter. This approach is, of course, very similar to and motivated

by the familiar Bode-Shannon[2] approach to optimum linear filtering for stationary processes.

Let us suppose that we decompose the white process y(t) and the process x(t) that we wish to represent into the orthonormal series

$$y(t) = \sum_{i=1}^{\infty} y_i \psi_i(t)$$
$$x(t) = \sum_{i=1}^{\infty} x_i \phi_i(t) \tag{50}$$

so that the random variables

$$y_1, y_2, y_3, \cdots$$
$$x_1, x_2, x_3, \cdots$$

represent the processes. Now, suppose we do this in such a way that

$$E[y_i x_j] = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases} \tag{51}$$

where $\lambda_1^2 \geqslant \lambda_2^2 \geqslant \ldots$. If we want to represent the set $\{x_i\}$ by n linear operations

$$z_j = \sum_{j=1}^{\infty} K_i y_i \qquad j = 1, \ldots, n$$

in such a way that the total mean-square error is minimum, intuitively we would first try to approximate the variable with the highest correlation, and then the next, and so on. For the approximation of x_1, we would minimize

$$E[(x_1 - z_1)^2] = E[x_1^2] - 2 \sum_{i=1}^{\infty} K_i E[x_1 y_i] + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} K_i K_j E[y_i y_j]$$
$$= E[x_1^2] - 2K_1 E[x_1 y_1] + \sum_{i=1}^{\infty} K_i^2.$$

Now

$$\frac{\partial}{\partial K_1} E[(x_1 - z_1)^2] = -2E[x_1 y_1] + 2K_1 = 0,$$

so that

$$K_1 = E[x_1 y_1] = \lambda_1$$

$$K_i = 0 \qquad i = 2, 3, \ldots .$$

The total error is then

$$\sum_{i=1}^{\infty} E[x_i^2] - E^2[x_1 y_1].$$

Thus we would approximate the set $\{x_i\}$ by

$$z_i = E[x_i y_i]\, y_i \qquad i = 1, \ldots, n.$$

Now the question is what orthonormal sets $\{\psi_i(t)\}$ and $\{\phi_i(t)\}$ do we use for Eqs. (49) so that conditions (50) hold? We want

$$E[x_i y_j] = \int_\Omega \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, \psi_j(t)\, ds\, dt = 0 \qquad i \neq j$$

Then we use the solutions[58]

$$\int_\Omega R_{xy}(s,t)\, \psi_i(t)\, dt = \lambda_i \phi_i(s) \qquad s \in \Omega$$

or

$$\int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds = \lambda_i \psi_i(t) \qquad t \in \Omega$$

for which

$$\int_\Omega \left[\int_\Omega R_{xy}(s,u)\, R_{xy}(t,u)\, du \right] \phi_i(t)\, dt = \lambda_i^2 \phi_i(s) \qquad s \in \Omega$$

Therefore, we use

$$z_i = a_i = \lambda_i \int_\Omega y(t)\, \psi_i(t)\, dt = \int_\Omega y(t)\, g_i(t)\, dt,$$

where

$$g_i(t) = \int_\Omega R_{xy}(s,t)\, \phi_i(s)\, ds$$

and

$$x(t) \approx z(t) = \sum_{i=1}^{n} a_i \phi_i(t)$$

in which the $\phi_i(t)$ are solutions of Eq. (52). The error is

$$\theta = \int_\Omega R_x(t,t)\, dt - \sum_{i=1}^{n} \lambda_i^2$$

and we are in agreement with our previous results in the first special case.

If the process y(t) is not white, we can make it so by performing the linear operation

$$y_1(s) = \int_\Omega R_y^{-1/2}(s,t)\, y(t)\, dt,$$

where $R_y^{-1/2}(s,t) = \sum_{i=1}^{\infty} \beta_i^{-1/2} e_i(s)\, e_i(t)$. The β_i and $e_i(t)$ are the eigenvalues and eigenfunctions of $R_y(s,t)$. To show that $y_1(s)$ is white, we take its autocorrelation function

$$R_{y_1}(s,t) = E[y_1(s)\, y_1(t)] = \int_\Omega \int_\Omega R_y^{1/2}(s,u)\, R_y^{-1/2}(t,v)\, R_y(u,v)\, du\, dv$$

$$= \int_\Omega R_y^{1/2}(s,u)\, du \int_\Omega \sum_{i=1}^{\infty} \beta_i^{-1/2} e_i(t)\, e_i(v) \sum_{j=1}^{\infty} \beta_j e_j(u)\, e_j(v)\, dv$$

$$= \int_\Omega R_y^{-1/2}(s,u)\, du \sum_{i=1}^{\infty} \beta_i^{1/2} e_i(t)\, e_i(u) = \sum_{i=1}^{\infty} e_i(s)\, e_i(t).$$

If we take any function f(t) of integrable square and perform the operation

$$\int_\Omega R_{y_1}(s,t)\, f(t)\, dt = \int_\Omega \sum_{i=1}^{\infty} e_i(s)\, e_i(t)\, f(t)\, dt = \sum_{i=1}^{\infty} e_i(s) \int_\Omega e_i(t)\, f(t)\, dt$$

$$= f(s)$$

then this implies that

$$R_{y_1}(s,t) = \sum_{i=1}^{\infty} e_i(s)\, e_i(t) = \delta(s-t),$$

which proves our assertion. We have lost nothing in performing this operation, since we may recover y(t) by operating with $R_y^{1/2}(s,t)$. We now apply the results obtained for white processes. The kernel K(s,t) becomes

$$K(s,t) = \int_\Omega R_{xy_1}(s,u)\, R_{xy_1}(t,u)\, du;$$

but

$$R_{xy_1}(s,t) = E\left[x(s) \int_\Omega R_y^{-1/2}(t,u)\, y(u)\, du\right]$$

$$= \int_\Omega R_y^{-1/2}(t,u)\, R_{xy}(s,u)\, du$$

so that

$$K(s,t) = \int_\Omega \int_\Omega R_y^{-1/2}(u,v)\, R_{xy}(s,v)\, dv \int_\Omega R_y^{-1/2}(u,w)\, R_{xy}(t,w)\, dw\, du$$

$$= \int_\Omega \int_\Omega R_{xy}(s,v)\, R_{xy}(t,w) \int_\Omega R_y^{-1/2}(u,v)\, R_y^{-1/2}(u,w)\, du\, dw\, dv$$

$$= \int_\Omega \int_\Omega R_{xy}(s,v)\, R_{xy}(t,w)\, R_y^{-1}(v,w)\, dv\, dw \qquad (53)$$

and

$$a_i = \int_\Omega h_i(t)\, y_1(t)\, dt = \int_\Omega g_i(t)\, y(t)\, dt,$$

where

$$h_i(t) = \int_\Omega R_{xy_1}(x,t)\, \gamma_i(s)\, ds$$

$$= \int_\Omega \int_\Omega R_y^{-1/2}(t,u)\, R_{xy}(s,u)\, \gamma_i(s)\, du\, ds,$$

in which the $\gamma_i(t)$ are the eigenfunctions of $K(s,t)$. Now,

$$a_i = \int_\Omega h_i(s)\, y_1(s)\, ds = \int_\Omega h_i(s) \int_\Omega R_y^{-1/2}(s,t)\, y(t)\, dt\, ds$$

so that

$$g_i(t) = \int_\Omega R_y^{-1/2}(s,t)\, h_i(s)\, ds$$

$$= \int_\Omega R_y^{-1/2}(s,t)\, ds \int_\Omega \int_\Omega R_y^{1/2}(s,u)\, R_{xy}(v,u)\, \gamma_i(v)\, du\, dv$$

$$= \int_\Omega \int_\Omega R_y^{-1}(t,u)\, R_{xy}(v,u)\, \gamma_i(v)\, dv. \qquad (54)$$

We then see that Eqs. (53) and (54) agree with Eqs. (48) and (49).

4.4 TIME-VARIANT LINEAR SYSTEMS

The problem considered in sections 4.2 and 4.3 can be interpreted in a slightly different manner that underlines its close relationship to optimum time-variant linear operations. As we have pointed out, the optimum time-variant linear operation on a process y(t) to approximate x(t) is given by

$$z(t) = \int_{\Omega} h(t, u)\, y(u)\, du \tag{55}$$

where h(t, u) is the solution of

$$\int_{\Omega} h(t, u)\, R_y(u, v)\, du = R_{xy}(t, v) \qquad t, v \in \Omega$$

If we assume that u is a parameter, the kernel h(t, u) can be expanded in the series

$$h(t, u) = h_u(t) = \sum_{i=1}^{\infty} g_i(u)\, \gamma_i(t) \tag{56}$$

where $\{\gamma_i(t)\}$ is orthonormal and

$$g_i(u) = \int_{\Omega} h(t, u)\, \gamma_i(t)\, dt.$$

If we substitute Eq. (56) in Eq. (55) and interchange the order of summation and integration, we obtain

$$z(t) = \sum_{i=1}^{\infty} \gamma_i(t) \int_{\Omega} y(u)\, g_i(u)\, du$$

so that

$$z(t) = \sum_{i=1}^{\infty} a_i \gamma_i(t),$$

where

$$a_i = \int_{\Omega} y(u)\, g_i(u)\, du.$$

We can then conclude that on the basis of the results of section 4.2 the finite series

$$\sum_{i=1}^{n} g_i(u)\ \gamma_i(u),$$

in which $\{g_i(t)\}$ and $\{\gamma_i(t)\}$ are solutions of Eqs. (44-47), approximates the linear operation (55) in a most rapidly convergent manner, in the sense that the mean-square error between z(t) and

$$z_n(t) = \sum_{i=1}^{n} \gamma_i(t) \int_{\Omega} y(u)\ g_i(u)\ du \tag{57}$$

is minimized for every n.

If we wished to perform a filtering of y(t) over all time, then we could do so by dividing the time axis into a series of intervals of the form $[\ell T, (\ell+1)T]$, where ℓ is any integer, and then perform the optimum operation indicated in Eq. (57) for each interval. If the processes are cyclostationary, that is

$$R_y(s,t) = R_y(s+T, t+T)$$

$$R_{xy}(s,t) = R_{xy}(s+T, t+T)$$

then the $\{\gamma_i(t)\}$ and $\{g_i(t)\}$ are the same for each interval. ("Cyclostationary" means that the ensemble statistics of the process vary periodically with time. This word was coined by W. R. Bennett.) The finite term approximation in Eq. (57) can then be realized in

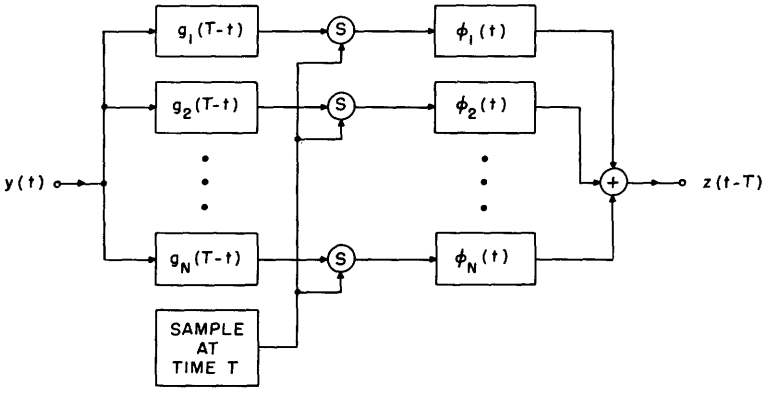


Fig. 11. The finite term approximation of a time-variant filter.

the form shown in Fig. 11. The process y(t) is first passed through filters of impulse responses $g_i(T-t)$, the outputs are then samples by an impulse at the end of each interval so that the result is an impulse of value a_i. The impulses then excite the second set of filters of impulse responses $\gamma_i(t)$, and the outputs are added together. The result is,

then, the filtered version $z_n(t-T)$ and we have a delay of T seconds. Thus, we have found an approximation of a time-variant filter for this cyclostationary case by using stationary components. The time-variant information comes from the knowledge of the sampling instants.

Single Time-Instant Estimation

Koschmann[21] has considered a problem that is related to the one considered here. It is the optimization of a set of coefficients $\{b_i\}$ in such a way that

$$E[(z(T_1)-x(T_1))^2] \tag{58}$$

is minimized, where $0 \leq T_1 \leq T$, and

$$z(T_1) = \sum_{i=1}^{\infty} b_i \int_0^T f_i(t)\, y(t)\, dt; \tag{59}$$

that is, the estimation of the value of the process x(t) at a single instant of time, which is based on an observation of the process y(t) during the whole interval. He showed that the optimum set $\{b_i\}$ must be a solution of the set of equations

$$\sum_{i=1}^{\infty} b_i \int\int_0^T R_y(u,v)\, f_i(u)\, f_j(v)\, du\, dv = \int_0^T R_{xy}(T_1, s)\, f_j(s)\, ds \qquad j = 1, 2, \ldots . \tag{60}$$

In order to show that our solution in which $b_i = \gamma_i(T_1)$ and $f_i(u) = g_i(u)$ are used also satisfies this condition, we substitute in Eq. (47) and, after inverting the order of integration, we obtain

$$\sum_{i=1}^{\infty} \gamma_i(T_1) \int_0^T \gamma_i(u)\, du \left\{ \int_0^T R_{xy}(u,v)\, g_j(v)\, dv \right\} = \int_0^T R_{xy}(T_1, s)\, g_j(s)\, ds.$$

The series on the left is an orthonormal series with Fourier coefficients and therefore it converges in the mean to the function on the right. Convergence in the mean insures that Eq. (60) is satisfied everywhere, except at points of a set of measure zero. See Courant and Hilbert.[59] Moreover, since our solution minimizes

$$\int_0^T E[(z_n(t)-x(t))^2]\, dt$$

where $z_n(t) = \sum_{i=1}^{n} \gamma_i(t) \int_0^T g_i(s)\, y(s)\, ds$, then we can say that although for our choice the series of Eq. (59) does not necessarily converge in a most rapid manner for

every T_1, it does so on the average over the interval.

4.5 WAVEFORM TRANSMISSION SYSTEMS

One example of a cyclostationary signal is the signal that occurs in a waveform transmission system. (An experimental system has been studied by Lovell, McGuigan, and Murphy.[25]) In such a system we have at the transmission end a set of n random variables $\{c_i\}$ that occur independently every T seconds. Each random variable multiplies one of an orthonormal set of waveforms $\{s_i(t)\}$, each of which is zero outside of the interval $[0, T]$, and the results are summed so that our resultant random waveform signal is

$$x(t) = \sum_{i=1}^{n} c_i s_i(t)$$

and

$$R_x(s, t) = \sum_{i=1}^{n} E[c_i^2]\, s_i(s)\, s_i(t) = \sum_{i=1}^{n} \lambda_i s_i(s)\, s_i(t),$$

where $\lambda_i = E[c_i^2]$. We shall assume that the signals are arranged in such a way that $\lambda_1 \geq \lambda_2 \geq \ldots$. If we transmit this signal over a noisy channel, we would then be interested in making an optimum linear estimation of the set $\{c_i\}$ based on the received signal y(t). We note that there is a difference between this problem and the one considered in sections 4.3 and 4.4. In this problem we are interested only in estimating the value of the parameters $\{c_i\}$, whereas before we were interested in estimating the entire wave shape.

Let us consider the case in which we want to find linear estimates $\{b_i\}$ of $\{c_i\}$, with $b_i = \int_0^T g_i(t)\, y(t)\, dt$, in such a way that

$$E\left[\sum_{i=1}^{n} (b_i - c_i)^2\right] \tag{61}$$

is minimized. This can be pictured by thinking of $\underline{c} = \{c_1, \ldots, c_n\}$ and $\underline{b} = \{b_1, \ldots, b_n\}$ as vectors. Then (61) is the average of the distance squared between the two vectors. This operation is equivalent to finding an estimate $z(t) = \sum_{i=1}^{n} b_i s_i(t)$ of x(t) in such a way that

$$E\left[\int_0^T (x(t) - z(t))^2\, dt\right]$$

is minimized, since

$$E\left[\int_0^T (x(t)-z(t))^2\,dt\right] = E\int_0^T \left[\sum_{i=1}^{n} c_i s_i(t) - \sum_{j=1}^{n} b_j s_j(t)\right]^2 dt$$

$$= E\left[\sum_{i=1}^{n} (b_i-c_i)^2\right].$$

We have already considered such a minimization in the first part of the representation problem considered in section 4.1, so that we see from Eq. (40) that $g_i(s)$ must satisfy

$$\int_0^T R_y(s,t)\, g_i(s)\, ds = \int_0^T R_{xy}(s,t)\, s_i(s)\, ds \qquad (62)$$

for $0 \leq t \leq T$.

The best linear estimates of the c_i are then realized by passing y(t) through filters of impulse responses $h_i(t) = g_i(T-t)$, and sampling at the end of each

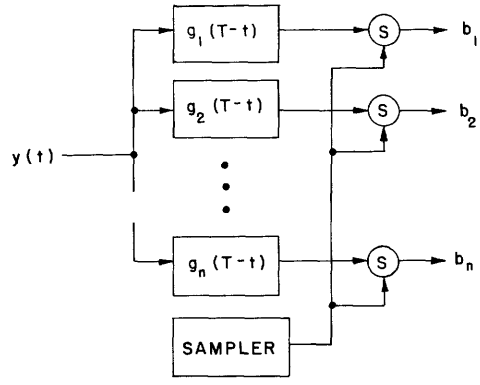


Fig. 12. The best linear estimator for the parameters c_i.

interval as shown in Fig. 12. If we have additive and independent noise, then

$$R_{xy}(s,t) = E[x(s)(x(t)+n(t))] = R_x(s,t)$$

so that Eq. (62) becomes

$$\int_0^T R_y(s,t)\, g_i(s)\, ds = \lambda_i s_i(t) \qquad 0 \leq t \leq T$$

which is the equation for the matched filter in the non-white noise case.[60] If the noise is white, that is $R_n(s,t) = N_o\delta(s-t)$, then we have

$$\int_0^T R_x(s,t)\, g_i(s)\, ds + N_o g_i(t) = \lambda_i s_i(t) \qquad 0 \leq t \leq T.$$

The solution is $g_i(s) = \frac{\lambda_i}{\lambda_i + N_o} s_i(s)$ for $i = 1, \ldots, n$, so that $h_i(t) = \frac{\lambda_i}{\lambda_i + N_o} s_i(T-t)$. In this case, by substituting in Eq. (43), the error becomes

$$\theta = \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \frac{\lambda_i^2}{\lambda_i + N_o} = \sum_{i=1}^{n} \frac{\lambda_i N_o}{\lambda_i + N_o}. \tag{63}$$

This linear estimator is a coherent device because its operation depends on the knowledge of the sampling instants; that is, any such system must include a method of extracting timing information from the signal.

4.6 WAVEFORM SIGNALS WITH MINIMUM BANDWIDTH

The problem with which we are concerned here has already been discussed by the author for a slightly different case.[15] When the random waveform signal is perturbed by independent white noise we see from Eq. (63) that the error is independent of the particular set of orthonormal waveforms which is used. We shall now concern ourselves with the problem of picking the set of waveforms in such a way that the expression

$$\int_{-\infty}^{\infty} f^2 S(f)\, df \tag{64}$$

is minimized, where S(f) is the power density spectrum of x(t). Expression (64) is the second moment of the spectrum and is, in a certain sense, a measure of the bandwidth. Of course, x(t) is not stationary so that it does not have a spectrum in the usual sense of the Wiener-Khinchin theorem. However, if we make the process stationary by assuming a random phase relationship between the members of the ensemble, we can then apply the Wiener-Khinchin theorem to the resulting stationary autocorrelation function. This is tantamount to using the time definition of the autocorrelation function with a single ensemble member

$$\mathscr{R}_x(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)\, dt.$$

In the following discussion we shall assume that the waveforms have continuous and bounded first derivatives and that $E[c_i] = 0$ for $i = 1, \ldots, n$ because if $E[c_i]$ were non-zero, periodicities would occur and S(f) would contain impulse functions. In such a case it would not be apparent what set of waveforms minimizes expression (64).

We can find S(f) by assuming that the random waveform signal was derived by applying impulse functions to a bank of linear filters with impulse responses $s_1(t), \ldots, s_n(t)$, which are zero for $t > T$, and adding the outputs, as shown in Fig. 13.

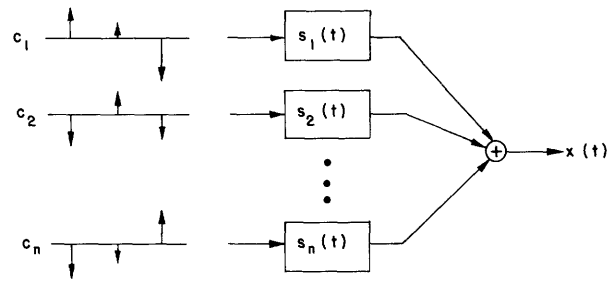


Fig. 13. Generation of a waveform signal.

The impulses are applied once every T seconds, and the impulse applied to the i^{th} filter has value c_i. Since the c_i are uncorrelated, the input processes are also. Letting $\delta(t)$ be the unit impulse function, we obtain for the individual input autocorrelation functions

$$R_i(t) = E[c_i^2]\,\delta(t) = \lambda_i \delta(t).$$

In accordance with the Wiener-Khinchin theorem the power density spectra are $\Phi_i(f) = \lambda_i$. It can be shown that the resulting output process has a power density spectrum[61]

$$S(f) = \sum_{i=1}^{n} |S_i(f)|^2\, \Phi_i(f) = \sum_{i=1}^{n} \lambda_i |S_i(f)|^2 \tag{65}$$

where

$$S_i(f) = \int_{-\infty}^{\infty} s_n(t) \exp(-j2\pi ft)\, dt = \int_0^T s_n(t) \exp(-j2\pi ft)\, dt. \tag{66}$$

This method to find S(f) is essentially the same as that used by Lee.[24]

Expression (66) now takes the form

$$\int_{-\infty}^{\infty} f^2 S(f)\, df = \sum_{i=1}^{n} \lambda_i \int_{-\infty}^{\infty} f^2 |S_i(f)|^2\, df$$

$$= \sum_{i=1}^{n} \lambda_i \int_{-\infty}^{\infty} f^2 S_i(f)\, \overline{S_i(f)}\, df \tag{67}$$

where the bar denotes the complex conjugate.

In order for the integral (64) to converge, it is necessary that

$$f^2 S(f) = 0(|f|^{-k})$$

for large f, with $k > 1$.[62] (Note that $f(x) = 0(g(x))$ signifies that $f(x)/g(x)$ remains bounded as x tends toward its limit.) Then

$$S(f) = 0(|f|^{-k-2}),$$

and from Eq. (67),

$$|S_i(f)|^2 = 0(|f|^{-k-2}) \qquad i = 1, 2, \ldots, n$$

or

$$|S_i(f)| = 0\left(|f|^{-\frac{k}{2}-1}\right) \qquad i = 1, 2, \ldots, n \tag{68}$$

where $k > 1$.

We shall now show that in order for Eq. (68) to hold, it is necessary that

$$s_i(0) = s_i(T) = 0 \qquad i = 1, 2, \ldots, n \tag{69}$$

Integrating Eq. (66) by parts, we get

$$S_i(f) = \frac{s_i(0) - s_i(T)\exp(-j2\pi ft)}{j2\pi f} + \frac{1}{j2\pi f}\int_0^T s_i'(t)\exp(-j2\pi ft)\,dt \tag{70}$$

in which the prime denotes differentiation.

Since the $s_i'(t)$ are bounded, $|s_i'(t)| < K$ for $0 \leq t \leq T$ for some number K. It follows that

$$\left|\frac{1}{j2\pi f}\int_0^T s_i'(t)\exp(-j2\pi ft)\,dt\right| \leq \frac{K}{|2\pi f|}\,\frac{|1-\exp(-j2\pi fT)|}{|2\pi f|}$$

$$= 0(|f|^{-2}).$$

Unless conditions (69) hold, it is seen that $S_i(t) = 0(|f|^{-1})$; and this violates Eq. (68). As seen from Eq. (70) with $s_i(0) = s_i(T) = 0$, the Fourier transforms of the $s_i'(t)$ are $(j2\pi f)\,S_i(f)$. From Parseval's theorem we obtain

$$\int_{-\infty}^{\infty}(j2\pi f)\,S_i(f)(-j2\pi f)\,\overline{S_i(f)}\,df = \int_{-\infty}^{\infty}[s_i'(t)]^2\,dt$$

$$= 4\pi^2\int_{-\infty}^{\infty} f^2|S_i(f)|^2\,df$$

so that from Eq. (67) we see that the minimization problem has reduced to the minimization of

$$\sum_{i=1}^{n}\lambda_i\int_0^T[s_i'(t)]^2\,dt \tag{71}$$

under the constraints that $\{s_i\}$ be an orthonormal set and $s_i(0) = s_i(T) = 0$ for all $i = 1, \ldots, n$. Integrating by parts, we obtain

$$\int_0^T [s_i'(t)]^2 \, dt = s_i'(T)\, s_i(T) - s_i'(0)\, s_i(0) - \int_0^T s_i(t)\, s_i''(t)\, dt$$

$$= - \int_0^T s_i(t)\, s_i''(t)\, dt$$

so that the minimization of (71) is equivalent to the maximization of

$$\sum_{i=1}^{n} \lambda_i \int_0^T s_i(t)\, s_i''(t)\, dt \tag{72}$$

which is

$$\sum_{i=1}^{n} \lambda_i \int_0^T s_i(t)\, L[s_i(t)]\, dt,$$

where L is the linear operator

$$L[f(t)] = \frac{d^2}{dt^2} f(t)$$

with boundary conditions $f(0) = f(T) = 0$.

This operator is self-adjoint, since

$$\int_0^T g(t)\, L[f(t)]\, dt = \int_0^T g(t)\, f''(t)\, dt = \int_0^T g'(t)\, f'(t)\, dt$$

$$= \int_0^T g''(t)\, f(t)\, dt = \int_0^T L[g(t)]\, f(t)\, dt$$

by integration by parts, where $g(0) = g(t) = 0$, and $f(0) = f(t) = 0$.

From Theorem I, expression (72) is then maximized by the first n solutions of

$$\frac{d^2}{dt^2} s_i(t) = \beta_i s_i(t)$$

with the boundary conditions $s_i(0) = s_i(T) = 0$. These solutions are

$$s_\ell(t) = \left[\frac{2}{T}\right]^{1/2} \sin \frac{\ell \pi}{T} t \qquad \begin{cases} 0 \leq t \leq T & \ell = 1, 2, \ldots \\ \text{elsewhere} \end{cases}$$

$$= 0$$

for which

$$\beta_\ell = -\left[\frac{\ell\pi}{T}\right]^2 \qquad \ell = 1, 2, \ldots .$$

For these solutions

$$|S_\ell(f)|^2 = \frac{8\ell^2 T}{(\ell^2 - 4f^2T^2)^2\,\pi^2}\cos^2 \pi fT \qquad \ell \text{ odd}$$

$$= \frac{8\ell^2 T}{(\ell^2 - 4f^2T^2)\,\pi^2}\sin^2 \pi fT \qquad \ell \text{ even}$$

From Eq. (65) the power density spectrum becomes

$$S(f) = \frac{8T}{\pi^2}(\cos \pi fT)^2 \sum_{\substack{\ell=1 \\ \ell \text{ odd}}}^{n} \frac{\lambda_\ell \ell^2}{(\ell^2 - 4f^2T^2)^2} + \frac{8T}{\pi^2}(\sin \pi fT)^2 \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{n} \frac{\lambda_\ell \ell^2}{(\ell^2 - 4f^2T^2)^2} .$$

The power density spectra obtained by using n = 1, 2, 3 are shown in Fig. 14

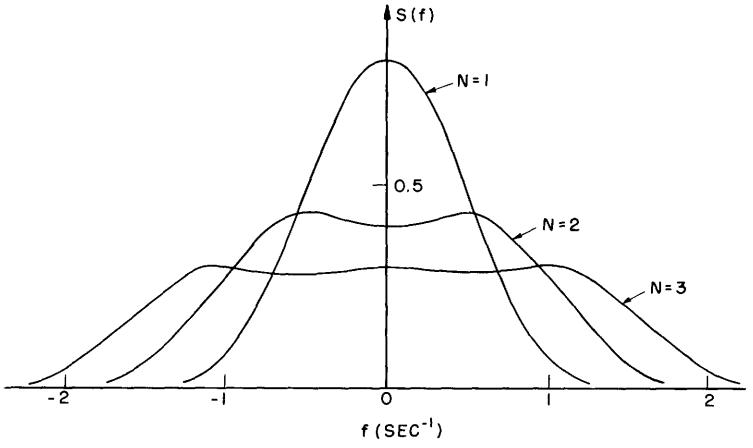


Fig. 14. Power density spectrum obtained by using the optimum waveforms.

for T = 1. In these examples it was assumed that $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \frac{1}{n}$.

Let us consider a normalized version of this spectrum,

$$S_N(f) = \frac{S\left[f\frac{n}{2T}\right]}{S(0)},$$

for the case in which $\lambda_1 \geq \lambda_2 \geq \ldots$. After some algebraic manipulation, we find that

$$S_N(f) = \frac{1}{\sum_{\substack{\ell=1 \\ \ell \text{ odd}}}^{n} \frac{1}{\ell^2}} \left[\left[\cos \frac{\pi f n}{2} \right]^2 \sum_{\substack{\ell=1 \\ \ell \text{ odd}}}^{n} \frac{\ell^2}{(\ell^2 - f^2 n^2)^2} + \left[\sin \frac{\pi f n}{2} \right]^2 \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{n} \frac{\ell^2}{(\ell^2 - f^2 n^2)^2} \right]$$

$$\leq \sum_{\ell=1}^{n} \frac{\ell^2}{(\ell^2 - f^2 n^2)^2} .$$

If $f > 1$, then

$$\frac{\ell^2}{(\ell^2 - f^2 n^2)^2} \leq \frac{n^2}{(n^2 - f^2 n^2)^2}$$

for $\ell \leq n$, so that for $f > 1$

$$\sum_{\ell=1}^{n} \frac{\ell^2}{(\ell^2 - f^2 n^2)^2} \leq n \frac{n^2}{(n^2 - f^2 n^2)^2} = \frac{1}{n(1-f^2)^2},$$

and therefore, we have the following upper bound for the spectrum for $f > 1$:

$$S_N(f) \leq \frac{1}{n(1-f^2)^2} .$$

This tells us that it is possible to make up a signal with waveforms that are time-limited to T seconds in such a way that the signal has, on the average, n/T degrees of freedom per unit time, and the power contained outside of a bandwidth of n/2T cps is vanishingly small for n that is large enough. We note in this respect that if we drop the time-limited restriction we can do it with zero power outside of a bandwidth of n/2T cps by using $\frac{\sin x}{x}$ functions.

V. THE NUMERICAL COMPUTATION OF EIGENFUNCTIONS

In general, the analytical solution of integral equations of the form

$$\int_{\Omega} K(s, t)\, \phi(t)\, dt = \lambda\phi(s) \qquad s \in \Omega \tag{73}$$

is a formidable problem. In view of this we have developed a computer program for the solution of these equations on the IBM 704 computer at the Computation Center, M.I.T. A general description of the methods used in this program follows.

5.1 THE COMPUTER PROGRAM

The program can be divided into three main sections:

(i) The approximation of the integral equation by a matrix equation.

(ii) The diagonalization of the matrix equation.

(iii) The manipulation of the diagonalizing matrix to obtain the desired approximation of the eigenfunctions.

For approximating the integral equation by a matrix equation, we use the Gauss-Legendre quadrature method for the approximation of a definite integral. Assuming that the integral has been normalized in such a manner that the interval of integration is $[-1, 1]$, we approximate the integral by a finite sum

$$\int_{-1}^{1} f(t)\, dt = \sum_{i=1}^{n} a_i f(t_i),$$

where the weights a_i and the abscissas t_i are to be chosen. If we specify that the approximation above be exact for $f(t) = 1, x, x^2, \ldots, x^{2n-1}$, then we have $2n$ equations and $2n$ unknowns, and we can solve for the a_i's and t_i's. The approximation is then exact for any polynomial of degree $2n - 1$ or less. The weights and abscissas are tabulated for the interval $[-1, 1]$ for n up to 16.[63] If a more accurate approximation is desired, the interval can be divided into subintervals with a separate approximation for each interval. In this program we have used a 10-point approximation for the basic interval, so that n will be any multiple of 10.

If we apply this method to Eq. (73), we obtain

$$\int_{\Omega} K(s, t)\, \phi(t)\, dt \approx \sum_{j=1}^{n} a_j K(s, t_j)\, \phi(t_j) = \lambda\phi(s)$$

and considering this for the same values of s as for t, we get the following set of linear equations:

$$\sum_{j=1}^{n} a_j K(t_i, t_j)\, \phi(t_j) = \lambda \phi(t_i) \qquad i = 1, \ldots, n.$$

We now make the substitution $\gamma(t_j) = \sqrt{a_j}\, \phi(t_j)$ from which we have

$$\sum_{j=1}^{n} \sqrt{a_i}\, K(t_i, t_j)\, \sqrt{a_j}\, \gamma(t_j) = \lambda \gamma(t_i) \qquad i = 1, \ldots, n.$$

These equations are now symmetrical, and can be solved by diagonalizing the matrix

$$[\sqrt{a_i}\, K(t_i, t_j) \sqrt{a_j}\,].$$

This is done by means of an efficient and accurate subprogram[64] written by F. J. Corbato of the Computation Center, M.I.T. This program gives the eigenvalues λ_k, and the diagonalizing matrix with the eigenvectors $\gamma_k(t_j)$ as columns.

Our approximations of the eigenfunctions $\phi_k(t)$ of Eq. (73) are then,

$$\phi_k(t_j) = \frac{1}{\sqrt{a_j}}\, \gamma_k(t_j) \qquad k, j = 1, \ldots, n.$$

We now have n samples of each of the approximations of the eigenfunctions. These samples are rather far apart, and in order to find intermediate values we have to interpolate. The interpolation is done separately for each subinterval by assuming that the function is a linear combination of the first ten Legendre functions $\ell_i(t)$:

$$\sum_{i=1}^{10} a_i \ell_i(t_j) = f(t_j) \qquad j = 1, \ldots, 10,$$

so that we have 10 equations and 10 unknowns, each equation corresponding to one sample point or abscissa. We then solve for the a_i's by using a program for solving linear equations (we used program No. ANF402).

The time required for the running of the program on the IBM 704 computer for n = 40 is approximately 10-15 minutes.

We have described the program operation for a finite interval of integration. If the interval is $\Omega = [0, \infty]$, we can approximate the integral equation in a similar fashion. In our program we have divided the time axis into the four subintervals $[0, 3]$, $[3, 8]$, $[8, 16]$, and $[16, \infty]$. In the first three we have used a 10-point Gauss-Legendre approximation, and in the last interval we have used a 15-point Gauss-Laguerre approximation so that we have a 45×45 matrix. The Gauss-Laguerre approximation is used when the integral to be approximated is over the semi-infinite interval and is similar to the Gauss-Legendre, except that it is specified that the approximation be exact for

$f(t) = e^{-t}, xe^{-t}, \ldots, x^{2n-1}e^{-t}$.[65] The remaining operations are then the same as before.

5.2 THE COMPARISON OF NUMERICAL RESULTS WITH A KNOWN SOLUTION

In order to check the accuracy of the program, we have used it to compute the solutions of an example the results of which are known analytically. We have used as a kernel

$$K(s, t) = \pi e^{-2\pi|s-t|}.$$

The eigenfunctions and eigenvalues for this kernel have been given in section 3.7 but are repeated here for convenience. The eigenfunctions are:

$$\phi_k(t) = \begin{cases} c_k \cos b_k t & k \text{ odd} \\ c_k \sin b_k t & k \text{ even} \end{cases}$$

where the c_k's are normalizing constants, and the b_k's are the solutions of the transcendental equations

$$b_k \tan b_k A = 2\pi \qquad k \text{ odd}$$

$$b_k \cot b_k A = -2\pi \qquad k \text{ even.}$$

The eigenvalues are given by

$$\lambda_k = \frac{4\pi^2}{b_k^2 + 4\pi^2}.$$

The transcendental equations were solved and the eigenvalues and eigenfunctions for k = 1, 2, 3, 6, and 10 were found to be:

$\lambda_1 = 0.7105$ $\qquad$ $\phi_1(t) = 0.830 \cos 1.003t$

$\lambda_2 = 0.3392$ $\qquad$ $\phi_2(t) = 0.907 \sin 2.193t$

$\lambda_3 = 0.1632$ $\qquad$ $\phi_3(t) = 0.952 \cos 3.558t$

$\lambda_6 = 0.0367$ $\qquad$ $\phi_6(t) = 0.989 \sin 8.047t$

$\lambda_{10} = 0.0120$ $\qquad$ $\phi_{10}(t) = 0.996 \sin 14.247t.$

The eigenvalues computed by the program for n = 20 were:

$\lambda_1 = 0.7136$

$\lambda_2 = 0.3426$

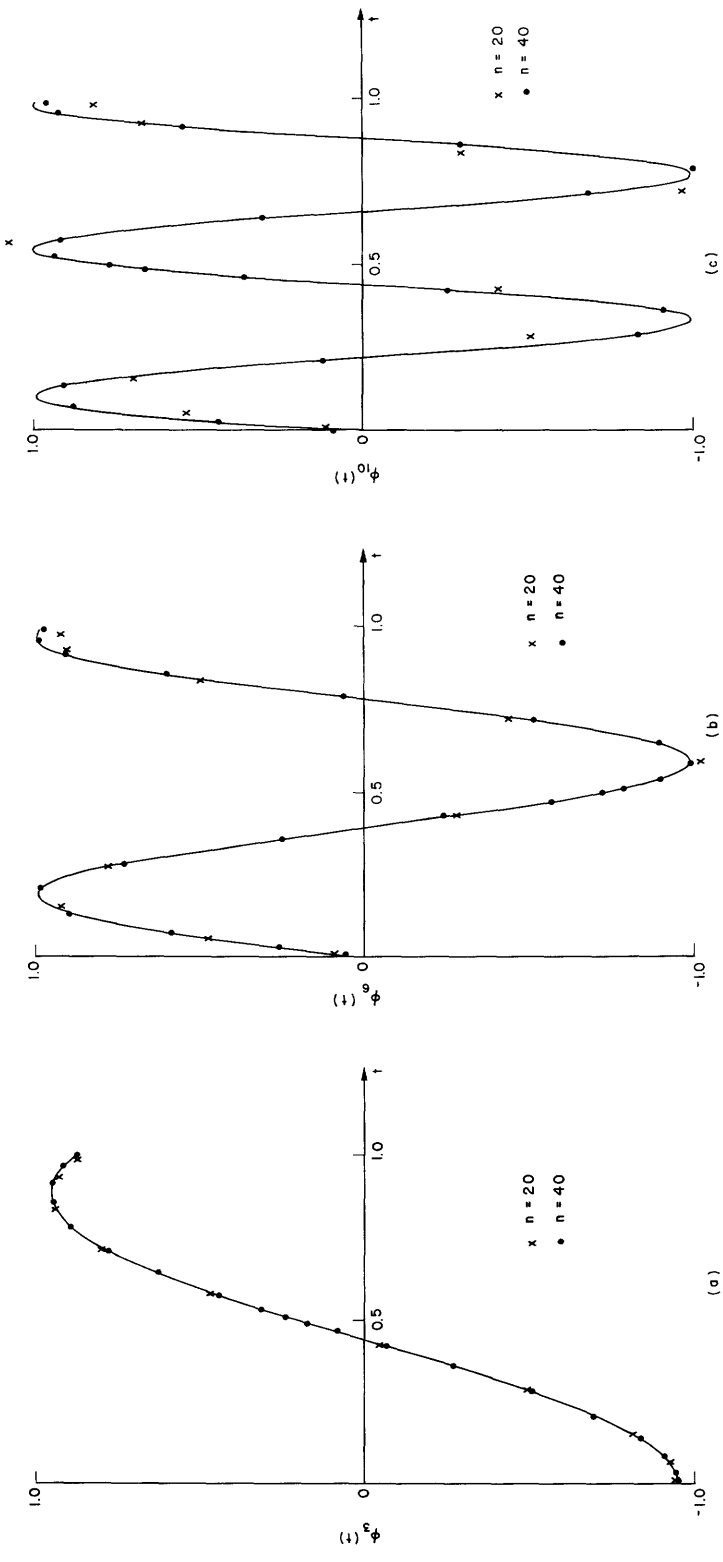


Fig. 15. Comparison of actual and computed forms of: (a) $\phi_3(t)$; (b) $\phi_6(t)$; (c) $\phi_{10}(t)$.

$$\lambda_3 = 0.1655$$

$$\lambda_6 = 0.0399$$

$$\lambda_{10} = 0.0160$$

and those for n = 40 were:

$$\lambda_1 = 0.7113$$

$$\lambda_2 = 0.3400$$

$$\lambda_3 = 0.1640$$

$$\lambda_6 = 0.0375$$

$$\lambda_{10} = 0.0128.$$

The sample points for the computer eigenfunctions over one-half of the interval are shown plotted with the true eigenfunctions in Fig. 15. The first two eigenfunctions $\phi_1(t)$ and $\phi_2(t)$ are not shown because there was no discernible difference between the actual and the computed.

5.3 THE EXPERIMENTAL COMPARISON OF EIGENFUNCTIONS AND LAGUERRE FUNCTIONS FOR THE EXPANSION OF THE PAST OF A PARTICULAR RANDOM PROCESS

The optimum set of functions for expanding the past of a signal can, in some cases, do much better than Laguerre functions. To show this we have taken a sample function of a random process generated in the laboratory and expanded it by means of the digital computer. We chose a zero-mean random process with correlation function

$$R(\tau) = \exp\left[-|\tau|\right] \cos 3\tau,$$

and used a weighted norm with weighting function $W(t) = \exp\left[-t/4\right]$. (We use the terminology of section 3.9.) The process has power density spectrum

$$S(f) = \frac{1}{1 + 4\pi^2\left[f + \frac{3}{2\pi}\right]^2} + \frac{1}{1 + 4\pi^2\left[f - \frac{3}{2\pi}\right]^2}.$$

The autocorrelation function and power density spectrum are shown in Figs. 16 and 17. Such a process was generated by passing white noise through a filter with system function

$$H(s) = \sqrt{2}\,\frac{s + \sqrt{10}}{s^2 + 2s + 10}$$

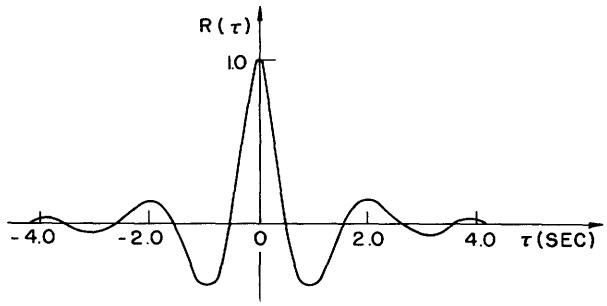


Fig. 16. Autocorrelation function of the process that is to be represented experimentally.

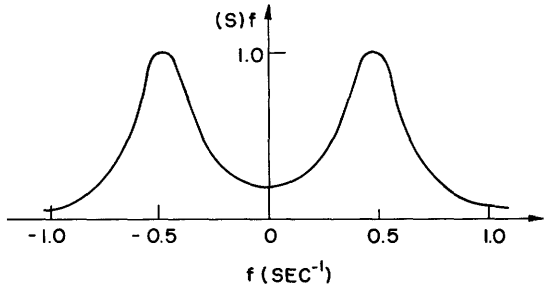


Fig. 17. Power density spectrum of the process that is to be represented experimentally.

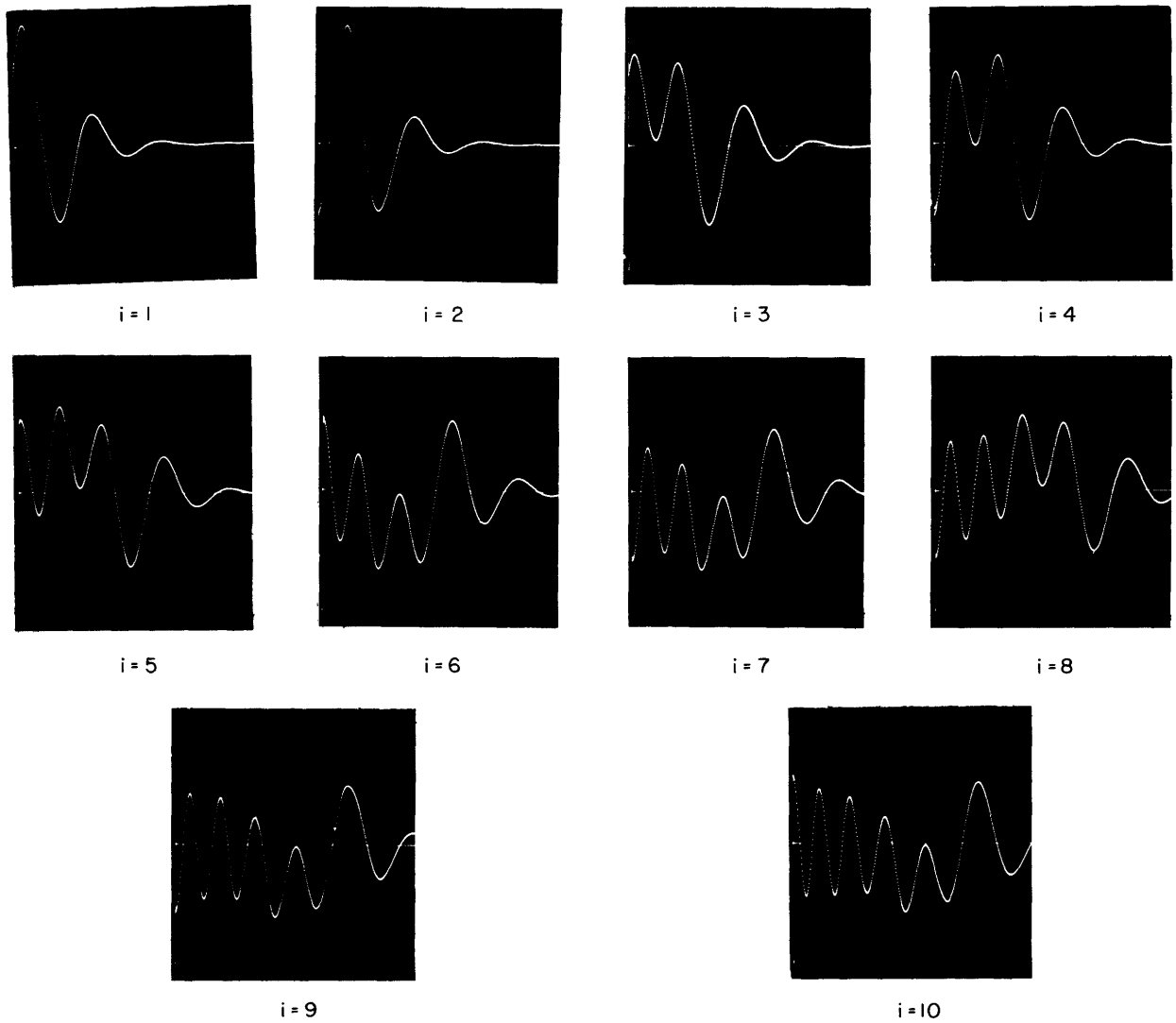
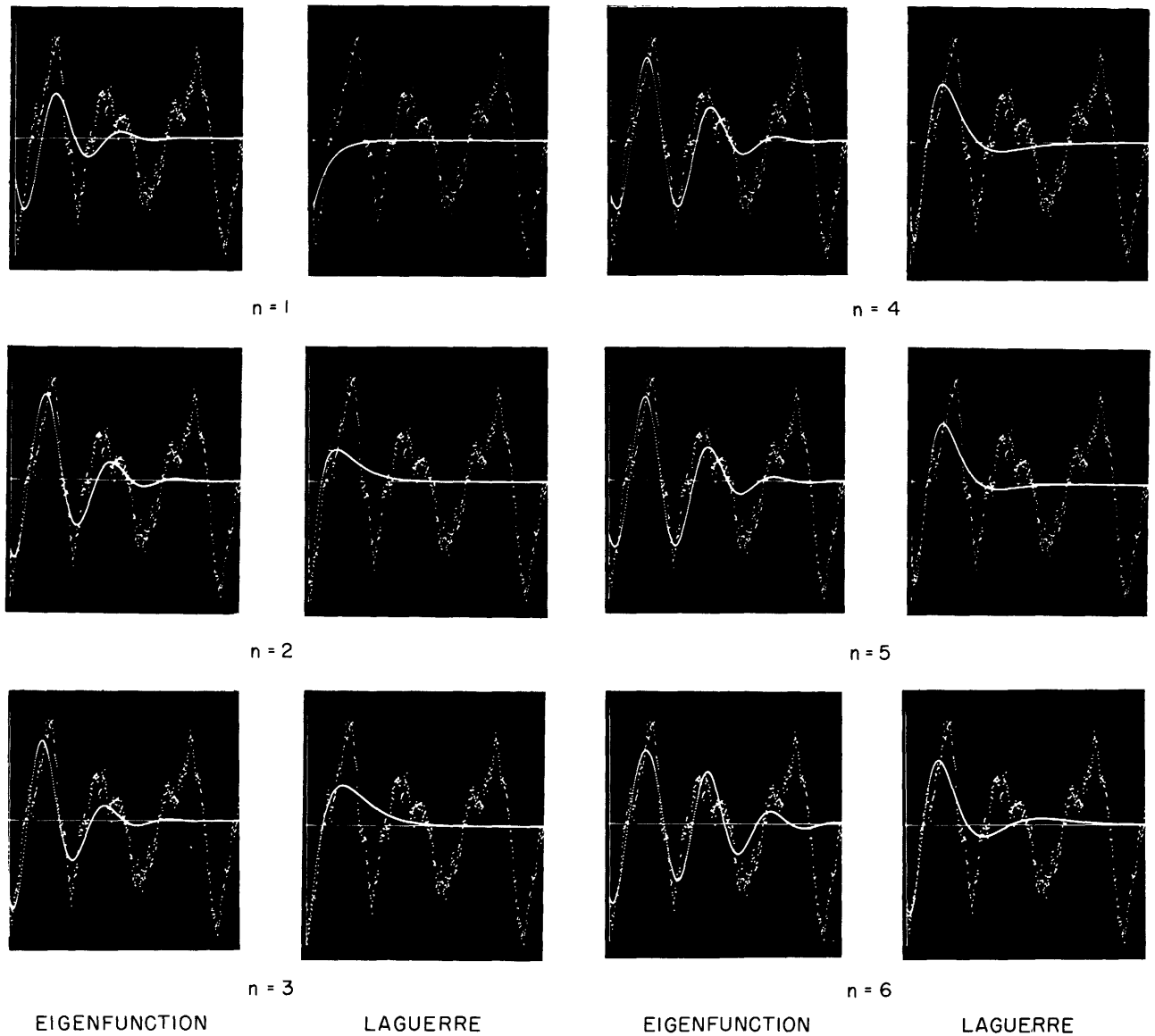


Fig. 18. Eigenfunctions of R(t) as computed on the IBM 704 computer.


n = 1
n = 4
n = 2
n = 5
n = 3
n = 6
EIGENFUNCTION
LAGUERRE
EIGENFUNCTION
LAGUERRE

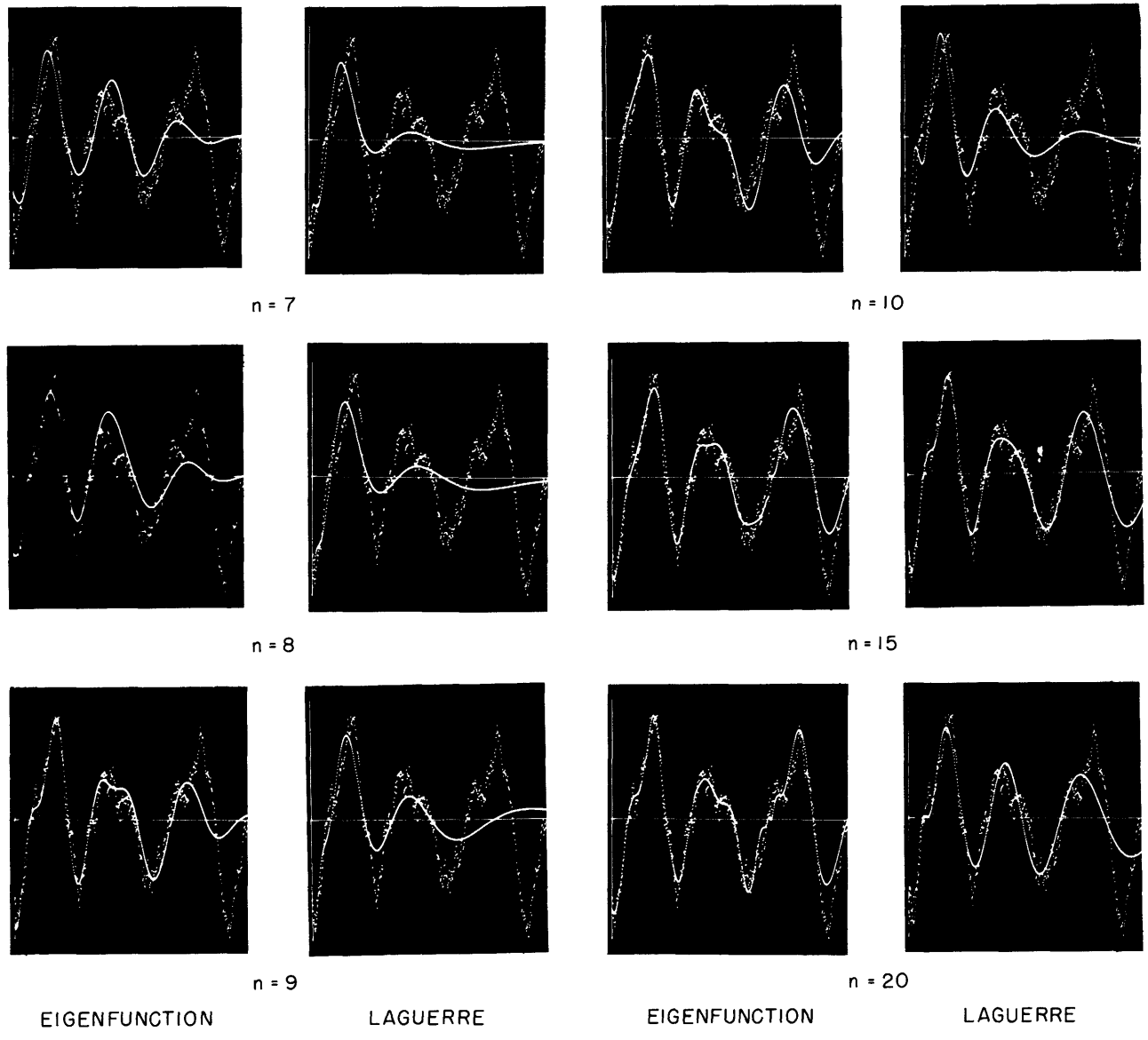


Fig. 19. Comparison of the approximation of a sample of the process with the use of eigenfunctions and Laguerre functions.

(Of course, the actual circuit that was used was scaled up in frequency and impedance level, but this is irrelevant here.)

The first ten eigenfunctions computed by the program for the integral equation

$$\int_0^\infty \exp\left[-\frac{s}{4}-\frac{t}{4}-|s-t|\right] \cos 3(s-t)\, \phi(t)\, dt = \lambda\phi(s).$$

are shown in Fig. 18.

The scale factor that was used for the Laguerre functions was chosen by minimizing the weighted error for the first Laguerre function in a manner similar to that used in section 3.10. The scale factor found on this basis was $\alpha = 4.5$.

The approximations of a sample function of the process over a period of 7.5 seconds by using the eigenfunctions and Laguerre functions in a straight orthogonal expansion for n = 1, ..., 10, 15, and 20 terms is shown in Fig. 19. It is seen that the eigenfunctions do much better, especially in approximating the higher frequency portions than the Laguerre functions. This is, as we pointed out in section 3.10, because the Laguerre functions have Fourier transforms of the form

$$\frac{1}{n!}\,\frac{\left(j2\pi f-\frac{1}{2}\right)^n}{\left(j2\pi f+\frac{1}{2}\right)^{n+1}} \qquad n = 0, 1, 2, \ldots$$

so that most of their energy is near the origin. As we see from Fig. 17, however, most of the energy in the random process is not near the origin, so that the performance of the Laguerre functions is not expected to be near optimum.

APPENDIX A

PROOF OF THEOREM I

THEOREM: The sum

$$\sum_{i=1}^{n} c_i \left\langle \phi_i, L[\phi_i] \right\rangle$$

with $c_1 \geq c_2 \geq \ldots \geq c_n$, is maximized with respect to the orthonormal set of functions $\{\phi_i(t)\}$ by the choice

$$\phi_i(t) = \gamma_i(t) \qquad i = 1, 2, \ldots, n$$

and this maximum value is

$$\sum_{i=1}^{n} c_i \lambda_i$$

PROOF: First, the eigenfunctions are the solutions of

$$L[\gamma_i(t)] = \lambda_i \gamma_i(t) \qquad i = 1, 2, \ldots$$

arranged so that $\lambda_1 \geq \lambda_2 \geq \ldots$. Since L is self-adjoint, the $\gamma_i(t)$ form an orthogonal set. If the $\gamma_i(t)$ are normalized, we see that

$$\sum_{i=1}^{n} c_i \left\langle \gamma_i, L[\gamma_i] \right\rangle = \sum_{i=1}^{n} c_i \int_\Omega \gamma_i(t)\, L[\gamma_i(t)]\, dt$$

$$= \sum_{i=1}^{n} c_i \lambda_i$$

Now we shall show that this is the maximum. Suppose we have some other orthonormal set $\{\phi_i(t)\}$ for which

$$\omega_{ij} = \left\langle \phi_i(t), \gamma_j(t) \right\rangle = \int_\Omega \phi_i(t)\, \gamma_j(t)\, dt$$

Then

$$\int_\Omega \phi_i(t)\, L[\phi_i(t)]\, dt = \int_\Omega \phi_i(t)\, L\left[\sum_{j=1}^{\infty} \omega_{ij} \gamma_j(t)\right] dt$$

$$= \int_\Omega \phi_i(t) \sum_{j=1}^{\infty} \lambda_j \omega_{ij} \gamma_j(t)\, dt = \sum_{j=1}^{\infty} \lambda_j \omega_{ij}^2$$

$$= \lambda_n \sum_{j=1}^{\infty} \omega_{ij}^2 + \sum_{j=1}^{n} (\lambda_j - \lambda_n)\, \omega_{ij}^2 + \sum_{j=n+1}^{\infty} (\lambda_j - \lambda_n)\, \omega_{ij}^2$$

$$\leq \lambda_n + \sum_{j=1}^{n} (\lambda_j - \lambda_n)\, \omega_{ij}^2$$

Since $\sum_{j=1}^{\infty} \omega_{ij}^2 = 1$, all i, and $\lambda_j - \lambda_n \leq 0$ all $j \geq n+1$, then

$$\sum_{i=1}^{n} \int_{\Omega} \phi_i(t)\, L[\phi_i(t)]\, dt \leq n\lambda_n + \sum_{j=1}^{n} (\lambda_j - \lambda_n) \sum_{i=1}^{n} \omega_{ij}^2$$

$$= n\lambda_n + \sum_{j=1}^{n} (\lambda_j - \lambda_n) + \sum_{j=1}^{n} (\lambda_j - \lambda_n) \left[\sum_{i=1}^{n} \omega_{ij}^2 - 1 \right]$$

Therefore,

$$\sum_{j=1}^{n} \lambda_j - \sum_{i=1}^{n} \int_{\Omega} \phi_i(t)\, L[\phi_i(t)]\, dt \leq \sum_{j=1}^{n} (\lambda_j - \lambda_n) \left[1 - \sum_{i=1}^{n} \omega_{ij}^2 \right]$$

Now, since $\lambda_j - \lambda_n \geq 0$, $j = 1, \ldots, n$, and

$$0 \leq \sum_{i=1}^{n} \omega_{ij}^2 \leq \sum_{i=1}^{\infty} \omega_{ij}^2 = 1$$

then

$$\sum_{i=1}^{n} \int_{\Omega} \phi_i(t)\, L[\phi_i(t)]\, dt \leq \sum_{j=1}^{n} \lambda_j$$

and this true for any n. Now consider

$$c_1\lambda_1 + c_2\lambda_2 + \ldots + c_n\lambda_n = c_n(\lambda_1 + \ldots + \lambda_n)$$
$$+ (c_{n-1} - c_n)(\lambda_1 + \ldots + \lambda_{n-1}) + \ldots +$$
$$+ (c_1 - c_2)(\lambda_1)$$

If we set $a_i = \int_{\Omega} \phi_i(t)\, L[\phi_i(t)]\, dt$, we know that

$$\lambda_1 + \ldots + \lambda_n \geq a_1 + \ldots + a_n$$

$$\lambda_1 + \ldots + \lambda_{n-1} \geq a_1 + \ldots + a_{n-1}$$
$$\vdots \qquad\qquad \vdots$$
$$\lambda_1 \geq a_1$$

If we multiply consecutively by c_n, $c_{n-1} - c_n$, ..., $c_1 - c_2 \geq 0$ and add, we get

$$\sum_{i=1}^{n} c_i \lambda_i \geq \sum_{i=1}^{n} c_i \alpha_i = \sum_{i=1}^{n} c_i \int_{\Omega} \phi_i(t)\, L[\phi_i(t)]\, dt$$

which was to be proved. The proof of the second case is similar.

APPENDIX B

PROOF OF THEOREM II.

We shall first state the following theorem of Kac, Murdock, and Szegö[18] which will be used to prove Theorem II.

THEOREM. Consider the integral equation

$$\int_{-A}^{A} \rho(s-t)\, \phi_i(t)\, dt = \lambda_i \phi_i(s) \qquad -A \leq s \leq A$$

with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. If we define

$$F(f) = \int_{-\infty}^{\infty} \rho(t) \exp(-j2\pi ft)\, dt \tag{B-1}$$

then

$$\lim_{A\to\infty} \frac{1}{2A} N_A(a, b) = \mu[f; a < F(f) < b] \tag{B-2}$$

Here, $N_A(a, b)$ is the number of eigenvalues of the integral equation having values falling within (a, b), and $\mu[E]$ denotes the measure (or length for our purposes) of the set E. The limit (B-2) is true provided that (a, b) does not contain zero, and the sets with $F(f) = a$ or $F(f) = b$ are of measure zero.

If $R_x(t) = \rho(t)$ in the theorem above, $F(f)$ is then $S_x(f)$, the power density spectrum of the process $x(t)$, and is therefore even and everywhere positive. Let us assume that $S_x(f)$ is continuous and monotonically decreasing for positive arguments. We then subdivide the interval (a_o, b_o) in the range S_x into n subintervals, denoting the subdivision by $(a_o, a_1, \ldots, a_{n-1}, a_n=b_o)$ where $a_o = S_x(f_o)$ and $b_o = S_x(0)$. The corresponding subdivision of the positive domain is $(f_n=0, f_{n-1}, \ldots, f_1, f_o)$, where $a_i = S_x(f_i)$. We now observe that from the theorem

$$2a_{i-1}(f_{i-1}-f_i) \leq \lim_{A\to\infty} \frac{1}{2A} \sum_{D_i} \lambda_\ell \leq 2a_i(f_{i-1}-f_i)$$

where $D_i = [\ell; a_{i-1} \leq \lambda_\ell \leq a_i]$, and from this it follows that

$$2 \sum_{i=1}^{n} a_{i-1}(f_{i-1}-f_i) \leq \lim_{A\to\infty} \frac{1}{2A} \sum_{D} \lambda_\ell \leq 2 \sum_{i=1}^{n} a_i(f_{i-1}-f_i)$$

where $D = [\ell; \lambda_\ell \geq a_o]$. This is true for any subdivision and, by the definition of the Riemann integral,[51] if $S_x(f)$ is integrable, then

$$\text{l.u.b. } 2\sum_{i=1}^{n} a_{i-1}(f_{i-1}-f_i) = \text{g.l.b. } 2\sum_{i=1}^{n} a_i(f_{i-1}-f_i)$$

$$= 2\int_0^{f_o} S_x(f)\,df = \int_{-f_o}^{f_o} S_x(f)\,df$$

and we have

$$\lim_{A\to\infty} \frac{1}{2A}\sum_{D} \lambda_i = \int_{-f_o}^{f_o} S_x(f)\,df$$

Placing $f_o = k/2$, we get

$$\lim_{A\to\infty} \frac{1}{2A}\sum_{D} \lambda_i = \int_{-k/2}^{k/2} S_x(f)\,df$$

where $D = [i;\lambda_i \geq S_x(k/2)]$. We then observe that from the theorem

$$\lim_{A\to\infty} \frac{1}{2A} N_A(S_x(k/2), \infty) = k$$

or $N_A(S_x(k/2), \infty) = n \sim 2kA$, so that we have finally

$$\lim_{n\to\infty} \frac{k}{n}\sum_{i=1}^{n} \lambda_i = \int_{-k/2}^{k/2} S_x(f)\,df$$

A similar result can be obtained for monotonic spectra subject to the conditions of the theorem. It amounts to adjusting a_o in such a way that

$$\mu[f; S_x(f) \geq a_o] = k$$

We then have

$$\lim_{n\to\infty} \frac{k}{n}\sum_{i=1}^{n} \lambda_i = \int_{E} S_x(f)\,df$$

where $E = [f; S_x(f) \geq a_o]$, and this result is used in section 3.5.

APPENDIX C

THE ERROR INCURRED BY SAMPLING AND RECONSTRUCTING BY MEANS OF $\frac{\sin x}{x}$ FUNCTIONS

Let x(t) be a random process with autocorrelation function R(t) and power density spectrum S(f). We sample the process at the rate $\frac{1}{2W}$ samples per second and reconstruct with $\frac{\sin x}{x}$ functions and obtain a new process,

$$y(t) = \sum_{n=-\infty}^{\infty} x(n\tau) \frac{\sin \frac{\pi}{\tau}(t-n\tau)}{\frac{\pi}{\tau}(t-n\tau)}$$

where $\tau = \frac{1}{2W}$. We want to find the error

$$\begin{aligned} E[e^2(t)] &= E[(y(t)-x(t))^2] \\ &= E[y^2(t)] + E[x^2(t)] - 2E[x(t)y(t)] \\ &= E[x^2(t)] + \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} R(n\tau-m\tau) \frac{\sin \frac{\pi}{\tau}(t-n\tau) \sin \frac{\pi}{\tau}(t-m\tau)}{\frac{\pi}{\tau}(t-n\tau)\frac{\pi}{\tau}(t-m\tau)} \\ &\quad - 2 \sum_{n=-\infty}^{\infty} R(t-n\tau) \frac{\sin \frac{\pi}{\tau}(t-n\tau)}{\frac{\pi}{\tau}(t-n\tau)} \end{aligned} \tag{C-1}$$

The second term may be reduced in the following manner (which is due to Slepian). Let $n - m = \ell$

$$\sum_{\ell=-\infty}^{\infty} R(\ell\tau) \sum_{m=-\infty}^{\infty} \frac{\sin \frac{\pi}{\tau}(t-\ell\tau-m\tau) \sin \frac{\pi}{\tau}(t-m\tau)}{\frac{\pi}{\tau}(t-\ell\tau-m\tau)\frac{\pi}{\tau}(t-m\tau)} \tag{C-2}$$

Since $\frac{\sin x}{x}$ is bandlimited, we have the identity

$$\frac{\sin \frac{\pi}{\tau}(x-a)}{\frac{\pi}{\tau}(x-a)} = \sum_{m=-\infty}^{\infty} \frac{\sin \frac{\pi}{\tau}(m\tau-a)}{\frac{\pi}{\tau}(m\tau-a)} \frac{\sin \frac{\pi}{\tau}(x-m\tau)}{\frac{\pi}{\tau}(x-m\tau)}$$

and letting $a = t$ and $\ell\tau = t - x$, we get

$$\frac{\sin \frac{\pi}{\tau}\ell\tau}{\frac{\pi}{\tau}\ell\tau} = \sum_{m=-\infty}^{\infty} \frac{\sin \frac{\pi}{\tau}(t-m\tau)}{\frac{\pi}{\tau}(t-m\tau)} \frac{\sin \frac{\pi}{\tau}(t-\ell\tau-m\tau)}{\frac{\pi}{\tau}(t-\ell\tau-m\tau)}$$

so that expression (C-2) becomes

$$\sum_{\ell=-\infty}^{\infty} R(\ell\tau)\,\frac{\sin\pi}{\pi} = R(0) = E[x^2(t)]$$

We now have for the series (C-1)

$$E[e^2(t)] = 2E[x^2(t)] - 2\sum_{n=-\infty}^{\infty} R(t-n\tau)\,\frac{\sin\frac{\pi}{\tau}(t-n\tau)}{\frac{\pi}{\tau}(t-n\tau)}$$

Now the last term is periodic of period τ, so we average over a period

$$\frac{1}{\tau}\int_{-\tau/2}^{\tau/2} E[e^2(t)]\,dt = 2E[x^2(t)] - \frac{2}{\tau}\int_{-\tau/2}^{\tau/2}\sum_{n=-\infty}^{\infty} R(t-n\tau)\,\frac{\sin\frac{\pi}{\tau}(t-n\tau)}{\frac{\pi}{\tau}(t-n\tau)}\,dt$$

$$= 2E[x^2(t)] - \frac{2}{\tau}\int_{-\infty}^{\infty} R(t)\,\frac{\sin\frac{\pi}{\tau}t}{\frac{\pi}{\tau}t}\,dt \qquad \text{(C-3)}$$

Here, we make use of the identity

$$\int_{-\tau/2}^{\tau/2}\sum_{n=-\infty}^{\infty} f(t-n\tau)\,dt = \int_{-\infty}^{\infty} f(t)\,dt$$

According to Parseval's theorem, we have

$$\int_{-\infty}^{\infty} R(t)\,\frac{\sin\frac{\pi}{\tau}t}{\frac{\pi}{\tau}t}\,dt = \tau\int_{-W}^{W} S_x(f)\,df$$

so that

$$\frac{1}{\tau}\int_{-\tau/2}^{\tau/2} E[e^2(t)]\,dt = 2E[x^2(t)] - 2\int_{-W}^{W} S_x(f)\,df$$

$$= 2\int_{-\infty}^{\infty} S_x(f)\,df - 2\int_{-W}^{W} S_x(f)\,df$$

$$= 2\left[\int_{-\infty}^{-W} S_x(f)\,df + \int_{W}^{\infty} S_x(f)\,df\right]$$

which was to be proved.

APPENDIX D

DETERMINATION OF THE EIGENVALUES AND EIGENFUNCTIONS OF A CERTAIN KERNEL

We shall find the eigenvalues and eigenfunctions of the kernel

$$K(s,t) = e^{-\beta s^2 - \beta t^2 - \frac{\alpha}{2}(s-t)^2} \tag{D-1}$$

First, we shall need two identities. The first is

$$\int_{-\infty}^{\infty} e^{-at^2} e^{-j2\pi ft}\, dt = \sqrt{\frac{\pi}{a}} \exp\left(-\frac{\pi^2}{a} f^2\right) \tag{D-2}$$

We have

$$\int_{-\infty}^{\infty} e^{-at^2} e^{-j2\pi ft}\, dt = \int_{-\infty}^{\infty} \exp\left[-a\left(t + j\frac{\pi}{a} f\right)^2\right] \exp\left(-\frac{\pi^2}{a} f^2\right) dt$$

$$= \exp\left(-\frac{\pi^2}{a} f^2\right) \int_{-\infty + j\frac{\pi}{a}f}^{\infty + j\frac{\pi}{a}f} e^{-at^2}\, dt$$

and since the integrand is entire,

$$= \exp\left(-\frac{\pi^2}{a} f^2\right) \int_{-\infty}^{\infty} e^{-at^2}\, dt = \sqrt{\frac{\pi}{a}} \exp\left(-\frac{\pi^2}{a} f^2\right)$$

The second identity is

$$\int_{-\infty}^{\infty} \left\{\frac{d^n}{dt^n} e^{-bt^2}\right\} e^{at^2} e^{-j2\pi ft}\, dt = \left[-\frac{ja}{\pi}\right]^n \sqrt{\frac{\pi}{b-a}} \exp\left(-\frac{\pi^2}{a} f^2\right) \left\{\frac{d^n}{df^n} \exp\left(-\frac{b\pi^2}{a(b-a)} f^2\right)\right\} \tag{D-3}$$

Consider

$$I = \int_{-\infty}^{\infty} \left\{\frac{d^n}{dt^n} e^{-bt^2}\right\} e^{at^2} e^{-jyt}\, dt$$

$$= e^{y^2/4a} \int_{-\infty}^{\infty} \left\{\frac{d^n}{dt^n} e^{-bt^2}\right\} \exp\left[a\left(t - j\frac{y}{2a}\right)^2\right]$$

Since the integrand is entire,

$$I = e^{y^2/4a} \int_{-\infty}^{\infty} \left\{ \frac{d^n}{dt^n} \exp\left[-b\left[t + j\frac{y}{2a} \right]^2 \right] \right\} e^{at^2}\, dt$$

$$= e^{y^2/4a} \left[\frac{2a}{j} \right]^n \frac{d^n}{dy^n} \int_{-\infty}^{\infty} \exp\left\{ -b\left[t + j\frac{y}{2a} \right]^2 \right\} e^{at^2}\, dt$$

$$= e^{y^2/4a} \left[\frac{2a}{j} \right]^n \frac{d^n}{dy^n} e^{by^2/4a^2} \int_{-\infty}^{\infty} \exp(-(b-a)t^2) \exp\left(-j\frac{byt}{a} \right) dt$$

$$= e^{y^2/4a} \left[\frac{2a}{j} \right]^n \frac{d^n}{dy^n} \exp\left(\frac{b}{4a^2} y^2 \right) \sqrt{\frac{\pi}{b-a}} \exp\left(-\frac{b^2 y^2}{4a^2(b-a)} \right)$$

$$= \left[\frac{2a}{j} \right]^n \sqrt{\frac{\pi}{b-a}} e^{y^2/4a} \frac{d^n}{dy^n} \exp\left(-\frac{b}{4a(b-a)} y^2 \right)$$

If we let $y = 2\pi f$, then we have the identity.

Now we want to show that

$$\phi_n(t) = e^{kt^2} \frac{d^n}{dt^n} e^{-2kt^2} \tag{D-4}$$

are solutions of

$$\int_{-\infty}^{\infty} e^{-\beta s^2 - \beta t^2 - \frac{\alpha}{2}(s-t)^2} \phi_n(t)\, dt = \lambda_n \phi_n(s) \tag{D-5}$$

for some scale factor k. Substituting (D-4) in the left side of (D-5), we have

$$I = e^{-\beta s^2} \int_{-\infty}^{\infty} \exp\left[-\frac{\alpha}{2}(s-t)^2 \right] e^{(k-\beta)t^2} \left\{ \frac{d^n}{dt^n} e^{-2kt^2} \right\} dt$$

but since

$$\int_{-\infty}^{\infty} f(s)\, g(t-s)\, ds = \int_{-\infty}^{\infty} e^{j2\pi ft}\, F(f)\, G(f)\, df$$

where F and G are the Fourier transforms of f and g, we have

$$I = e^{-\beta s^2} \int_{-\infty}^{\infty} e^{j2\pi fs} \left[\sqrt{\frac{2\pi}{\alpha}} \exp\left(-\frac{2\pi^2}{\alpha} f^2 \right) \right]$$

$$\left[-\frac{j(k-\beta)}{\pi} \right]^n \sqrt{\frac{\pi}{k+\beta}} \exp\left(\frac{\pi^2}{k-\beta} f^2 \right) \left\{ \frac{d^n}{df^n} \exp\left(-\frac{2k\pi^2}{k^2 - \beta^2} f^2 \right) \right\} df$$

$$= \left[-\frac{j(k-\beta)}{\pi}\right]^n \sqrt{\frac{2\pi^2}{a(k+\beta)}}\, e^{-\beta s^2} \int_{-\infty}^{\infty} e^{j2\pi fs} \exp\left\{\left[\frac{\pi^2}{k-\beta} - \frac{2\pi^2}{a}\right] f^2\right\}$$

$$\left\{\frac{d^n}{df^n} \exp\left(-\frac{2k\pi^2}{k^2-\beta^2} f^2\right)\right\} \, df$$

Here, we have applied the identity (D-3). If we apply it again and simplify, we get

$$I = \left[1-\frac{2}{a}(k-\beta)\right]^n \sqrt{\frac{2\pi}{a+2(k+\beta)}} \exp\left\{\left[\frac{1}{\frac{1}{k-\beta}-\frac{2}{a}} - \beta\right] s^2\right\}$$

$$\left\{\frac{d^n}{ds^n}\, e^{-\frac{2kas^2}{\left[\frac{1}{k-\beta}-\frac{2}{a}\right](ka-\beta a+2(k^2-\beta^2))}}\right\}$$

If we set $k = \sqrt{\beta(a+\beta)}$, after some manipulation we see that

$$I = \sqrt{\frac{2\pi}{a+2\sqrt{\beta(a+\beta)}+2\beta}} \left[1-\frac{2}{a}(\sqrt{\beta(a+\beta)}-\beta)\right]^n e^{ks^2} \frac{d^n}{ds^n} e^{-2ks^2}$$

so that the eigenvalues and eigenfunctions are

$$\lambda_n = \sqrt{\frac{2\pi}{a+2\sqrt{\beta(a+\beta)}+2\beta}} \left[1-\frac{2}{a}(\sqrt{\beta(a+\beta)}-\beta)\right]^n$$

$$\phi_n(t) = A_n e^{\sqrt{\beta(a+\beta)}\, t^2} \frac{d^n}{dt^n} e^{-2\sqrt{\beta(a+\beta)}\, t^2}$$

for $n = 0, 1, 2, \ldots$. In our case $a = \beta = 1$ and the kernel was multiplied by $\sqrt{\frac{\pi}{2}}$, so that

$$\lambda_n = \pi \sqrt{\frac{1}{3+2\sqrt{2}}} (3-2\sqrt{2})^n$$

$$\phi_n(t) = A_n e^{\sqrt{2}\, t^2} \frac{d^n}{dt^n} e^{-2\sqrt{2}\, t^2}$$

for $n = 0, 1, 2, \ldots$.

APPENDIX E

THE SOLUTION OF A CERTAIN INTEGRAL EQUATION

We want to find the $\gamma_n(t)$ that solves the equation

$$A_n e^{\sqrt{2}\,t^2} \frac{d^n}{dt^n} e^{-2\sqrt{2}\,t^2} = e^{-t^2} \int_{-\infty}^{\infty} e^{-(t-u)^2} \gamma_n(u)\, du$$

$$A_n e^{(\sqrt{2}+1)t^2} \frac{d^n}{dt^n} e^{-2\sqrt{2}\,t^2} = \int_{-\infty}^{\infty} e^{-(t-u)^2} \gamma_n(u)\, du$$

If we take the Fourier transform of both sides, using the identity (D-3), we get

$$A_n\left[-\frac{j(\sqrt{2}+1)}{\pi}\right]^n \sqrt{\frac{\pi}{\sqrt{2}-1}}\, e^{\frac{\pi^2}{\sqrt{2}+1} f^2} \left\{\frac{d^n}{df^n} e^{-2\sqrt{2}\pi^2 f^2}\right\} = \sqrt{\pi}\, e^{-\pi^2 f^2}\, \Gamma_n(f)$$

where $\Gamma_n(f)$ is the Fourier transform of $\gamma_n(t)$. We then see that

$$\Gamma_n(f) = A_n\left[-\frac{j(\sqrt{2}+1)}{\pi}\right]^n \sqrt{\frac{1}{\sqrt{2}-1}}\, e^{\frac{\sqrt{2}}{\sqrt{2}+1}\pi^2 f^2} \frac{d^n}{df^n} e^{-2\sqrt{2}\pi^2 f^2}$$

Taking the inverse Fourier transform of both sides and simplifying, we get

$$\gamma_n(t) = A_n(\sqrt{2}+2)^n \frac{1}{\sqrt{\pi(2-\sqrt{2})}}\, e^{\frac{1+\sqrt{2}}{2+\sqrt{2}} t^2} \frac{d^n}{dt^n} e^{-\sqrt{2}\,t^2}$$

Acknowledgment

The author takes pleasure in expressing his appreciation to Professor Y. W. Lee, who supervised his thesis project and who has been a constant source of encouragement; to Professor A. G. Bose for his original suggestion and continuing interest; and to the staff of the Computation Center, M.I.T., for their cooperation. The author is also indebted to his fellow-students David A. Chesler, C. E. Wernlein, Jr., and Harry L. Van Trees, Jr., for their assistance on numerous problems; and to Professor C. E. Shannon for his pertinent suggestion regarding section 3.5 of this report.

References

1. A. V. Balakrishnan, A note on the sampling principle for continuous signals, Trans. IRE Vol. IT-3, pp. 10-21, 1957.
2. H. W. Bode and C. E. Shannon, A simplified derivation of least squares smoothing and prediction theory, Proc. IRE 38, 417-426 (1950).
3. R. C. Booton, An optimization theory for time-varying linear systems with non-stationary statistical inputs, Proc. IRE 40, 977-981 (1952).
4. A. G. Bose, A theory of nonlinear systems, Technical Report 309, Research Laboratory of Electronics, M.I.T., May 15, 1956.
5. J. L. Brown, Jr., Mean square truncation error in series expansions of random functions, J. Soc. Indust. Appl. Math. 8, 28-32 (1960).
6. R. Courant, Differential and Integral Calculus, Vol. 1 (Interscience Publishers, Inc., New York, 2d Edition, 1937).
7. R. Courant and D. Hilbert, Methods of Mathematical Physics (Interscience Publishers, Inc., New York, 1959).
8. W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Publishing Company, New York, 1958).
9. K. Fan, On a theorem of Weyl concerning eigenvalues of linear transformations, Proc. Nat. Acad. Sci. 35, 652-655 (1949).
10. B. Friedman, Principles and Techniques of Applied Mathematics (John Wiley and Sons, Inc., New York, 1956).
11. W. H. Huggins, Signal theory, Trans. IRE Vol. CT-3, pp. 210-216, 1956; The Use of Orthogonalized Exponentials, AFCRC Report No. TR-57-357, Johns Hopkins University, September 30, 1957.
12. E. Jahnke and F. Emde, Tables of Functions with Formulae and Curves (Dover Publications, New York, reprint 4th Edition, 1945).
13. K. L. Jordan, Jr., Minimization of truncation error in series expansions of random processes, Quarterly Progress Report No. 52, Research Laboratory of Electronics, M.I.T., January 15, 1959, pp. 70-78.
14. K. L. Jordan, Jr., Characterization of random signals with an application to non-stationary linear systems, Quarterly Progress Report No. 56, Research Laboratory of Electronics, M.I.T., January 15, 1960, pp. 117-125.
15. K. L. Jordan, Jr., Waveform signals with minimum spectral width, Quarterly Progress Report No. 57, Research Laboratory of Electronics, M.I.T., April 15, 1960, pp. 89-93.
16. K. L. Jordan, Jr., The asymptotic behavior of the error in a finite Karhunen-Loeve expansion of a stationary random process, Quarterly Progress Report No. 58, Research Laboratory of Electronics, M.I.T., July 15, 1960.
17. M. L. Juncosa, An integral equation related to Bessel functions, Duke Math. J. 12, 465-471 (1945).
18. M. Kac, W. L. Murdock, and G. Szegö, On the eigenvalues of certain Hermitian forms, J. Rational Mechanics and Analysis 2, 767-800 (1953); cf. Theorem 2.
19. K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Ac. Sci. Fennicae, AI 37, Helsinki (1947).
20. K. Karhunen, Über die Struktur Stationärer zufälliger Funktionen, Ark. Mat. 1, 141-160 (1949).
21. A. Koschmann, On the filtering of nonstationary time series, Ph.D. thesis, Purdue University, August 1954, pp. 66-70; Proc. National Electronics Conference, 1954, p. 126, cf. Theorem 2.

22. H. P. Kramer, On best approximation of random processes et al., Trans. IRE Vol. IT-6, pp. 52-53, 1960.

23. Y. W. Lee, Synthesis of electric networks by means of the Fourier transforms of Laguerre's functions, J. Math. Phys. 2, 83-113 (1932).

24. Y. W. Lee, Autocorrelation functions of Poisson waves, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., January 15, 1955, pp. 56-62.

25. C. A. Lovell, J. H. McGuigan, and O. J. Murphy, An experimental polytonic signalling system, Bell System Tech. J. 34, 783-806 (1955).

26. I. G. Petrovskii, Lectures on the Theory of Integral Equations (translated from Russian by H. Kamel and H. Komm) (Graylock Press, Rochester, N. Y., 1957).

27. W. Rudin, Principles of Mathematical Analysis (McGraw-Hill Publishing Company, New York, 1953).

28. C. E. Shannon, Communications in the presence of noise, Proc. IRE 37, 10-21 (1949).

29. F. Smithies, Integral Equations, Cambridge Tracts in Mathematics and Mathematical Physics No. 49 (Cambridge University Press, London, 1958).

30. D. W. Tufts, A sampling theorem for stationary random processes, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1958, pp. 111-116; Quarterly Progress Report No. 53, April 15, 1959, pp. 87-93.

31. Tables of Functions and Zeros of Functions, National Bureau of Standards Applied Mathematics Series, Vol. 37 (U. S. Government Printing Office, Washington, D.C., 1954).

32. R. Courant, op. cit., p. 386.

33. B. Friedman, op. cit., pp. 18-22.

34. R. Courant and D. Hilbert, op. cit., p. 122.

35. I. G. Petrovskii, op. cit., p. 61

36. F. Smithies, op. cit., p. 127.

37. I. G. Petrovskii, op. cit., p. 72.

38. Ibid., p. 76.

39. Ibid., p. 68.

40. F. Smithies, op. cit., p. 131.

41. R. Courant and D. Hilbert, op. cit., p. 160

42. B. Friedman, op. cit., pp. 110-113; 125-127.

43. Ibid., p. 96.

44. K. Fan, op. cit., see Theorem 1.

45. F. Smithies, op. cit., p. 134.

46. W. B. Davenport and W. L. Root, op. cit., p. 96.

47. For a definition of the Stieltjes integral, see W. Rudin, op. cit., p. 88.

48. R. Courant and D. Hilbert, op. cit., p. 50.

49. I. G. Petrovskii, op. cit., cf. Theorem, p. 18.

50. W. Rudin, op. cit., cf. Theorem 3.14, p. 41.

51. Ibid., p. 88.

52. W. B. Davenport and W. L. Root, op. cit., pp. 99-101.

53. R. Courant and D. Hilbert, op. cit., p. 184.

54. E. Jahnke and F. Emde, op. cit., p. 166.

55. See R. Courant and D. Hilbert, op. cit., p. 98. The usual definition of Hermite functions includes a $(-1)^i$, but we have neglected this.

56. Ibid., p. 95.

57. R. C. Booton, op. cit., cf. Eq. 24; our equation is slightly different because we confine ourselves to filtering that is based on a finite time interval.

58. R. Courant and D. Hilbert, op. cit., p. 159.

59. Ibid., p. 110.

60. W. B. Davenport and W. L. Root, op. cit.; pp. 244-247; see especially Eq. (11-47).

61. Ibid., p. 114.

62. R. Courant, op. cit., p. 250.

63. For a brief description of the method and a tabulation of these values, see Tables of Functions and Zeros of Functions, op. cit., pp. 185-189.

64. SHARE Identification No. MIDH13.

65. For a description and tabulation, see Tables of Functions and Zeros of Functions, op. cit., pp. 191-199.