INFORMATION AND CONTROL 8, 390–422 (1965)

On the Measurement and Use of Time-Varying Communication Channels*

WILLIAM L. ROOT

The University of Michigan, College of Engineering, Instrumentation Engineering, Ann Arbor, Michigan

In radio, radar, sonar, and seismic signal detection there is often the problem of processing received signals which have been distorted by a linear operation in the process of being transmitted. Examples are scattering and multiple-path propagation of radio waves. Usually the nature of this linear operation cannot be known very precisely in advance, and it often is changing in time, so that in order to carry out effective processing of the received signals it is necessary repeatedly to test and measure the mode of transmission, or channel as it will be called.

In this paper the beginnings of a theory are established concerning time-varying and random linear channels with the intent of characterizing classes of channels which can be determined exactly or approximately by measurement, showing how the measurements can be made, analyzing the errors, and applying the results to the theory of signal detection.

The notion of a determinable class of channels is defined and general examples are given. These include classes of channels that are time-invariant, periodic, and which vary with a known trend. The measurement of slowly-varying channels by approximation by time-invariant ones belonging to a known determinable class is discussed. Relation between almost-time-invariance of a channel and the correlation properties of a kind of stationary random channel are developed and tied-in with the channel measurement theory. An application is made to the problem of detecting sure signals in noise when the channel is slowly-varying.

---

* This paper was originally prepared at the Mathematics Research Center, United States Army, the University of Wisconsin under Contract No. DA-11-022-ORD-2059 and revised at The University of Michigan under National Aeronautics and Space Administration grant Ns G-2-59.

I. INTRODUCTION

A considerable amount of work in recent years has gone into the study of how to process received radio, radar, sonar or seismic signals so as effectively to recover certain intelligence from these signals when they have been disfigured in transmission. Multiple ray paths, as occur in radio wave reflection from the ionosphere, or sound wave transmission in relatively shallow water, scattering from an irregular surface, such as the moon, or from randomly occurring inhomogeneities in the transmission medium, are typical phenomena which can result in time-varying, frequency-shifted and sometimes apparently random superpositions of the emitted waveform at the receiver. Usually, in addition, there is noise at the receiver of a highly random character of essentially thermal origin. A great many problems of signal processing in situations of the kind indicated can be based on a mathematical model in which the total received waveform is represented as the sum of two waveforms, one the result of a linear operation on the emitted signal, and the other a completely independent random noise. The linear operation may or may not be time-varying and may or may not be random. There may be unknown parameters determining the linear operation, the emitted signal, and the noise.

In this paper an attempt is made to begin a systematic study of certain aspects of the measurement and data processing problems arising when the linear operation on the signal (henceforth called the channel operation) is initially unknown. The primary concern is with measuring channel characteristics so that these characteristics may be available for communication signal processing. The special case of slowly-varying channels is considered in most detail.

We suppose that the total received signal $w(t)$ for both measurement and communication situations is of the form

$$w(t) = y(t;\alpha) + n(t), \qquad \tau_1 \leqq t \leqq \tau_2 \tag{1}$$

where $n(t)$ is noise and $y(t;\alpha)$ is the response of the linear channel to an input signal. In particular, we write

$$y(t;\alpha) = \int_{\Lambda(t)} h(t,s)x(s;\alpha)\,ds, \qquad \tau_1 \leqq t \leqq \tau_2 \tag{2}$$

where $x(t;\ \alpha)$ is for each α a known function of s representing the emitted signal, $h(t;s)$ is a kernel characterizing the channel and $y(t;\alpha)$

is, as in Eq. (1), the intelligence-bearing signal at the receiver. We are modeling the channel as a linear integral operator, or more properly as a collection of linear integral operators, depending upon τ_1, τ_2 and the sets $\Lambda(t)$, each with kernel $h(t, s)$ where $h(t, s)$ is presumed to be defined for $-\infty < t, s < \infty$. Thus, the channel is identified by the kernel $h(t, s)$. We shall sometimes take h to be an ordinary real-valued function, and sometimes take it to be a sample function from a stochastic process, i.e., $h(\cdot, \cdot) \equiv h(\cdot, \cdot, \omega)$ where ω is an element of a probability space. In the latter case we talk about stochastic channel operators.

In Section II, the basic definitions and notations are introduced. In Section III, the question is studied of how much prior information is needed about a channel in order that it can be precisely determined from measurement. This question is stated in the form: how can classes of possible channels be characterized so that a channel belonging to such a known class is identifiable from certain kinds of measurements? The formal definition of a determinable class is introduced as an answer to this question and examples are given. These examples include classes of time invariant channels, classes of periodic channels, and channels with known trend.

In Section IV, the measurement of slowly-varying channels is considered; the idea used is to approximate a slowly-varying channel by a time-invariant one belonging to a known determinable class. Error bounds are established. In Section V, the results of Section IV are applied to a study of the errors resulting in a classical sure-signal-in-noise detection problem when the channel is slowly varying.

The previous work which seems closest in spirit to most of this is that of Kailath (1959) on channel measurement. Superficially the approach here is different, more abstract and more general but with results less applicable from an engineering point of view. Not only the definition of a determinable class, but the idea of making such a definition and using it as a starting point is apparently new. It is hoped eventually to obtain information-theory-like results about channel measurement and use centered around the notion of determinable classes, but very little has been accomplished. In this connection it may be noted that bounded closed determinable classes are compact (this and other mathematical properties of determinable classes are shown in a forthcoming report by R. Prosser and the author) and therefore the notions of ϵ-entropy and ϵ-capacity (see Kolmogorov and Tihomorov (1959)) are applicable. For general background on time-varying channels see Price and Green (1960) and the survey paper by Kailath (1963) with its accompanying bibliography.

II. DEFINITIONS AND CONDITIONS

If $x(t)$ is the signal emitted during a time interval of interest, $a \leqq t \leqq b$, and $y(t)$ is the channel output during a time interval $c \leqq t \leqq d$, resulting from $x(t)$, we write, as in Eq. (2)

$$y(t) = \int_a^b h(t, s)x(s)\, ds, \qquad c \leqq t \leqq d. \tag{3}$$

Usually, but not always, $a = c$, $b = d$ (we shall from this point on consistently neglect a fixed minimum time of transmission). We shall always require $x(t)$ to be a real-valued measurable function, square-integrable on $[a, b]$.

The channel is characterized by the kernel $h(t, s)$ which usually is to be defined for $-\infty < t, s < \infty$, although occasionally it will be defined only for s, t in some suitable interval I. The channel is *deterministic* if $h(t, s)$ is a real-valued function; in this case it is required that h satisfy

$$\iint_\Lambda |h(t, s)|^2\, dt\, ds < \infty \tag{4}$$

for any bounded measurable set Λ in the plane. The channel is *stochastic* if

$$h(t, s) \equiv h(t, s; \omega)$$

is a real-valued stochastic process, with ω an element of a probability space Ω (the probability variable ω will be suppressed). If the channel is stochastic it is required that h satisfy

$$\iint_\Lambda E\,|h(t, s)|^2\, dt\, ds < \infty \tag{5}$$

for any bounded measurable set Λ. The condition (5) implies that (4) holds with probability one.

If we put $k(t, t - s) \equiv h(t, s)$ then the equation

$$y(t) = \int_a^b h(t, s)x(s)\, ds, \qquad a \leqq t \leqq b \tag{6}$$

may be written,

$$y(t) = \int_{t-b}^{t-a} k(t, u)x(t - u)\, du, \qquad a \leqq t \leqq b. \tag{7}$$

We shall refer to t, s, and u in these equations, respectively, as the *observation time*, *emission time*, and *age* variables. If the channel is deterministic the integral operator defined by Eq. (6) or (7) as an operator on $L_2(a, b)$ is Hilbert-Schmidt, for any finite a, b. If the channel is stochastic, then with probability one this operator is Hilbert-Schmidt. It is convenient to work with both forms of the kernel, and we shall continue to use the letters h and k as in Eqs. (6) and (7).

We say a deterministic channel is *realizable* if $k(t, u) = 0$ for all $u < 0$; has *finite memory* if there exists $\gamma(t) \geqq 0$ and bounded on every finite interval such that $k(t, u) = 0$ for all $u > \gamma(t)$, $-\infty < t < \infty$; is *time-invariant* if $k(t, u) = k(t', u)$ for all t, t', so that k does not actually depend on t. We shall say a stochastic channel is *realizable*, has *finite memory*, or is *time invariant* if for every finite t-interval $[a, b]$ the respective conditions above hold except on a set of sample functions of $k(t, s)$ of probability zero. (This implies of course that the conditions hold for all $k(\cdot, \cdot; \omega)$, $-\infty < s, t < \infty$, except for $\omega \in \Omega_0$, where prob $\Omega_0 = 0$.)

We shall assume in what follows that any stochastic channel to be considered will have the properties

$$Ek(t, u) = 0,$$

and

$$Ek(t, u)k(t', u') \equiv R(t, t'; u, u') \tag{8}$$

exists and is continuous in all its variables simultaneously. The first condition entails no loss of generality, because if there is a deterministic component it may be subtracted out and treated separately. The second condition will automatically imply (5). We can now define a stochastic channel to be *stationary in the observation time*[1] (ot-stationary) if $R(t, t'; u, u')$ is a function of t and t' only through their difference

[1] We do not bother to distinguish weak stationarity from strict stationarity, for, except in the Gaussian case where the two are the same, we are always concerned here with the former. See Bello (1962) for a complete classification of stochastic channels.

$t - t'$. In this case we write

$$Ek(t, u)k(t', u') \equiv R(t, t'; u, u') \equiv R(t - t'; u, u').$$

Before proceeding further the following notational conventions are established. If $f(t)$ is a square integrable function on $L_2[I]$, where $L_2[I]$ is the L_2-space with respect to Lebesgue measure on an interval I, we denote its norm by $\| f \|$ or $\| f \|_2$. If it is absolutely integrable we denote its norm in $L_1[I]$ by $\| f \|_1$. If $h(t, s)$ is square-integrable on $I \times I$, we denote its norm in $L_2[I \times I]$ by $[\![h]\!]$; $h(t, s)$ may then be the kernel of a Hilbert-Schmidt (HS) operator H on $L_2[I]$ and the HS norm of H is denoted also by $[\![H]\!]$. One has $[\![H]\!] = [\![h]\!]$. The usual operator norm is denoted by $\| H \|$.

We shall say a deterministic channel is *admissible* if it satisfies condition (4) and is realizable; a stochastic channel is admissible if it satisfies conditions (8) and is realizable. We note that for an admissible channel, with respect to the interval $[0, T]$,

$$[\![H]\!]^2 = \int_0^T \int_0^T | h(t, s) |^2 \, dt \, ds = \int_0^T \int_0^t | k(t, u) |^2 \, du \, dt.$$

If the channel operator H has a time-invariant kernel so that $k(t, u)$ does not actually depend on t, we put $g(u) \equiv k(t, u)$.

III. CHANNEL MEASUREMENTS AND DETERMINABLE CLASSES

Part of the over-all problem of communicating through unknown time-varying channels is making the short-term measurements which are intended to provide the temporarily valid estimates of channel behavior. There are questions of when these can be made and how. If there is no *a priori* information restricting the class of possible kernel functions $h(t, s)$, $0 \leqq t, s \leqq T$, then there is no way to determine $h(t, s)$, $0 \leqq t, s \leqq T$, by measurements performed during the observation interval $[0, T]$; i.e., given the equation $y = Hx$, $x \in L_2[0, T]$, H an arbitrary Hilbert-Schmidt operator on $L_2[0, T)$, there is no way to choose x so that knowledge of y determines H. Hence the class of possible kernels must be restricted in advance in such a way that for suitable x, $y = Hx$ does determine (or nearly determine) H.

A definition is stated below which is intended to offer a reasonable criterion as to when a class of channels can be measured effectively. The definition essentially imposes two kinds of restrictions: the first is to impose constraints to cut down on the "degrees of freedom" of H so that the equation $y = Hx$ can be solved uniquely for H, the second is to insure that H can be approximated arbitrarily closely with a finite set of measurements. The necessity of the first kind of restriction is evident if one considers the analogous situation (actually, a special case) in which x and y are n-vectors and H is an $n \times n$ matrix. For then, solving $y = Hx$ for H amounts to solving n equations for n^2 unknowns, unless additional information about H is available.

It should be noted that the point of view adopted in this section does not include the notion of any statistical characterization of the channel. For the moment at least the channel is treated as an unknown operator, not a random operator.

We now introduce precise definitions. By a *linear measurement* of a channel in the time interval $[0, T]$ is meant a finite collection of inner products (p_k, w), $k = 1, \cdots K$, $p_k \in L_2[0, T]$, defined when $w \in L_2[0, T]$, where

$$w(t) = [Hx](t) + n(t), \qquad 0 \leqq t \leqq T$$

is the received waveform, as in Eq. (1). In this context the transmitted signal $x(t)$ will sometimes be referred to as the test signal. In the definition to follow $n(t) = 0$.

We shall say a class $\mathcal{H}$ of admissible channel operators H is *uniformly determinable* (ϵ, I) if in the time interval I there is a test signal $x(t)$, a linear measurement $\{(p_1, w), \cdots, (p_k, w)\}$, and a function f from k-dimensional Euclidean space R_k to the HS operators on $L_2[I]$ which is continuous with respect to operator norm, such that for each H in the class $\mathcal{H}$

$$\hat{H} = f((p_1, Hx), \cdots, (p_k, Hx)) \tag{9}$$

is an admissible operator and

$$\| H - \hat{H} \| \leqq \epsilon.$$

The test signal $x(t)$, the linear measurement, and the function f we call a channel *determination* (ϵ, I). If for fixed I there is for each $\epsilon > 0$ a determination (ϵ, I), we say the class of channels is *uniformly determinable* $(0, I)$. If for each $\epsilon > 0$ there is an interval $I(\epsilon)$, where $I(\epsilon)$ approaches ∞ as $\epsilon \rightarrow 0$ and a determination $(\epsilon, I(\epsilon))$, we say the class of channels is uniformly determinable $(0, \infty)$.

The notion of uniform determinability is not restricted to classes of time-invariant channels, as we show later by examples, but we consider them first. They are of importance here especially as approximations for slowly-varying channels and as prototypes for channels with known trend. First, we observe that the class of all admissible time-invariant kernels is *not* uniformly determinable (ϵ, I), where I is any finite interval and $\epsilon > 0$ is quite arbitrary. In fact, consider any (ϵ, I) determination of H, $f((\phi_1, Hx), \cdots, (\phi_k, Hx))$. Now

$$(Hx)(t) = \int_0^t g(u)x(t-u)\,du, \qquad 0 \leqq t \leqq T$$

can be interpreted as an operator X with kernel $x(t-u)$ operating on $g \in L_2[0, T]$.

Hence the mapping ψ carrying g into z defined by

$$z = \psi(g) = ((\phi_1, Hx), \cdots, (\phi_k, Hx)) = ((\phi_1, g*x), \cdots, (\phi_k, g*x))$$

is a bounded linear mapping from $L_2[0, T]$ into a finite-dimensional linear space, and cannot be 1:1. Let $g' \neq g''$ and suppose $\psi g' = \psi g''$. Then for any constant $a > 0$, $\psi(ag' - ag'') = 0$, so that the kernels ag' and ag'' will yield the same determination, while $\| ag' - ag'' \|$ may be as large as desired. Then if H' and H'' are the convolution operators with kernels g' and g'' respectively, $\| aH' - aH'' \|$ may be made as large as desired, thus violating the assertion that there was given an (ϵ, I) determination.

If one considers the restricted class of admissible time-invariant kernels for which $\| H \| < C =$ constant, then a trivial refinement of the above argument shows that for each I there is an $\epsilon_0 > 0$ such that this class is not uniformly determinable (ϵ, I) for $\epsilon < \epsilon_0$.

Example 1. There are various ways of putting further restrictions on the class of admissible time-invariant kernels to make them uniformly determinable. For example, suppose $\{\phi_k\}$ is a complete orthonormal set in $L_2[0, T]$, then the class of all admissible time-invariant kernels $g(u)$ whose Fourier coefficients with respect to the ϕ_k are dominated in magnitude by the elements of a fixed sequence belonging to l_2 is uniformly determinable $(0, I)$, where I is the interval $[0, T]$.

To prove this we consider determinations in which the test signal $x(t)$ is an approximate δ-function and the function f is given by a partial sum of a Fourier series. Let H be the unknown channel operator with kernel $k(t, u) \equiv g(u)$ as before. We note that since $x(t)$ and $g(t)$ both vanish for $t \leqq 0$,

$$[Hx](t) = \int_0^t g(u)x(t-u)\,du$$
$$= \int_{-\infty}^{\infty} x(t-u)g(u)\,du, \qquad 0 \leqq t \leqq T.$$

Take $x_n(t)$ to be an approximate identity in L_2 under convolution such that $x_n(t)$ vanishes outside $[0, T]$ (e.g., $x_n(t) = n$ for $0 \leqq t \leqq 1/n$, zero otherwise, satisfies this condition, but there is a wide choice of such x_n, including many sequences of continuous functions). Then the $L_1(0, T)$-norm of x_n, $\| x_n \|_1$, is equal to one for all n ($\| x_n \|$ must approach ∞), and if

$$y_n(t) = \int_{-\infty}^{\infty} x_n(t-u)g(u)\,du, \qquad 0 \leqq t \leqq T$$

then

$$\| y_n \| \leqq \| x_n \|_1 \, \| g \| = \| g \|$$

(this follows, *a fortiori* for this truncated convolution from the usual inequality for convolutions with $g(u)$ set equal to zero for $u > T$). We designate the truncated convolution above by $x_n * g$. The determinations referred to can now be written

$$\hat{g}_{K,n} = \sum_{k=1}^{K} (\phi_k , y_n)\phi_k = \sum_{k=1}^{K} (\phi_k , x_n * g)\phi_k$$

where K and n are positive integers.

Let $\{a_k\}$ be any sequence of real numbers such that $\sum_1^{\infty} a_k^2 < \infty$. Put

$$\sum_1^{\infty} a_k^2 = M.$$

We now consider the class of all kernels $g(u) = \sum_1^{\infty} b_k\phi_k(u)$, $0 \leqq u \leqq T$, for which $|b_k|^2 \leqq a_k^2$. Then,

$$\hat{g}_{K.n} = \sum_{k=1}^{K} (\phi_k , x_n * \sum_{p=1}^{\infty} b_p\phi_p)\phi_k .$$

Any determination of the kind in question is given by

$$\sum_{k=1}^{K} (\phi_k , y_n)\phi_k = \sum_{k=1}^{K} (\phi_k , x_n * g)\phi_k = \sum_{k=1}^{K} (\phi_k , x_n * \sum_{p=1}^{\infty} b_p\phi_p)\phi_k$$

for some positive integer K and some approximate δ-function x_n. Given an arbitrary $\epsilon > 0$, let K be chosen large enough so that $\sum_{K+1}^{\infty} a_k^2 < \epsilon^2 M^2$. Then let $N \geqq K$ be large enough that $\sum_{N+1}^{\infty} a_k^2 < \epsilon^2 M^2/K$, and in the sequence of approximate δ-functions $\{x_n\}$, let n be large enough that $\| \phi_k - x_n * \phi_k \| < \epsilon/N$ for $k = 1, \cdots, N$. Since x_n has L_1-norm of 1.

$$| (\phi_k , x_n * \phi_k) | \leqq \| \phi_k \| \ \| x_n \|_1 \| \phi_k \| = 1,$$

and from the condition on x_n it follows immediately that

$$| (\phi_k , x_n * \phi_k) | \geqq 1 - \epsilon/N, \qquad k = 1, \cdots, K,$$

and

$$| (\phi_j , x_n * \phi_k) | \leqq \epsilon/N, \qquad j \neq k, k = 1, \cdots, K.$$

Then,

$$\left\| g - \sum_{k=1}^{K} (\phi_k , y_n) \phi_k \right\| \leqq \left\| \sum_{k=1}^{K} [b_k - (\phi_k , y_n)] \phi_k \right\| + \left\| \sum_{K+1}^{\infty} b_k \phi_k \right\|. \quad (10)$$

By the choice of K the second term is dominated by $\epsilon^2 M^2$. For the coefficients in the first term one has,

$$b_k - (\phi_k , y_n) = b_k[1 - (\phi_k , x_n * \phi_k)] - \sum_{p \neq k}^{N} b_p (\phi_k , x_n * \phi_p) - \left(\phi_k , x_n \sum_{N+1}^{\infty} b_p \phi_p \right)$$

and hence,

$$\begin{aligned} | b_k - (\phi_k , y_n) | &\leqq | b_k | \frac{\epsilon}{N} + \sum_{p \neq k}^{N} | b_p | \frac{\epsilon}{N} + \| \phi_k \| \ \| x_n \|_1 \left\| \sum_{N+1}^{\infty} b_p \phi_p \right\| \\ &\leqq | b_k | \frac{\epsilon}{N} + \frac{M\epsilon}{\sqrt{N}} + \frac{M\epsilon}{\sqrt{K}}. \end{aligned}$$

Thus the square of the first term on the right side of the inequality (10) is bounded by

$$\begin{aligned} \sum_{k=1}^{K} | b_k - (\phi_k , y_n) |^2 &\leqq 4 \left\{ \frac{\epsilon^2}{N^2} - \sum_{k=1}^{K} | b_k |^2 + \frac{KM^2\epsilon^2}{N} + M^2\epsilon^2 \right\} \\ &\leqq 4M^2\epsilon^2 \left(\frac{1}{N^2} + \frac{K}{N} + 1 \right) \\ &\leqq 12M^2\epsilon^2, \end{aligned}$$

and we have

$$\left\| g - \sum_{k=1}^{K} (\phi_k, y_n)\phi_k \right\| \leqq 5M\epsilon$$

for all g satisfying the stated condition. This implies the corresponding error in HS operator norms is less than $5M\epsilon\sqrt{T}$.

If the $\phi_k(t)$ are taken to be the sines ane cosines of the ordinary Fourier series, this condition says it is sufficient for uniform determinability that the energies in each frequency component be uniformly bounded and tail off uniformly at high frequencies.

Example 2. Consider admissible kernels with the periodicity property $h(t, s) = h(t + T_0, s + T_0)$ for all real t, u, which have finite memory γ. Let n be an integer large enough that $nT_0 > \gamma$, and let $\{\phi_k(t)\}$ be a complete orthonormal set on $[0, nT_0]$. Each admissible operator, being Hilbert-Schmidt, can be expressed as an infinite matrix with coefficients

$$h_{kj} = \int_0^{nT_0} \int_0^{nT_0} h(t, s)\phi_k(t)\phi_j(s)\, dt\, ds$$

where

$$\sum_{k=1,j=1}^{\infty} h_{kj}^2 < \infty.$$

Let $\{a_{kj}\}$ be an infinite sequence of real numbers such that

$$\sum_{k=1,j=1}^{\infty} a_{kj}^2 < \infty.$$

Consider the subclass of the periodic kernels with finite memory which satisfy the condition $h_{kj}^2 \leqq a_{kj}^2$, $k, j = 1, 2, \cdots$. This subclass of periodic channels is uniformly determinable $[0, \infty)$. The proof runs parallel to the one in the preceeding example and will not be given. The idea is that by using $\phi_1(t)$ as a test signal an arbitrarily good approximation can be obtained in the time interval $[0, nT_0]$ for the first column of the matrix. After a relaxation interval of length nT_0, a second determination will yield an arbitrarily good approximation to the second column of the same matrix, etc. Channels with periodicity of this sort do not seem at the moment to be of very much practical interest in communication. However, a slight modification may be of interest. If the channel is a linear system ("plant" in control engineering) which is under man's control and can be reset to a fixed initial state after being probed, then it can be tested again as indicated and it will be uniformly determinable if the regularity conditions stated above are satisfied.

In practice one is presumably not really interested in knowing how a channel transmits all signals of finite energy, but only those in a certain subclass, as for example, those in a certain frequency band. The notion of determinability is extended, therefore, to apply to subclasses (not necessarily linear) of signal functions. A class of admissible channels C is *uniformly determinable* (ϵ, I) *with respect to* S, S a subset of $L_2[I]$, if there is a determination yielding a bounded linear operator $\hat{H}$ on $L_2[I]$ such that for each H belonging to C

$$\sup_{x \in S} \frac{\| Hx - \hat{H}x \|}{\| x \|} < \epsilon.$$

Example 3. An obvious and often practical way to get an approximate determination of a time-invariant channel is to estimate its transfer function. If the time-invariant kernel is $g(u)$, the transfer function is defined to be

$$G(f) = \int_{-\infty}^{\infty} g(u) e^{i2\pi f u} \, du.$$

It is assumed the channel has finite memory $\leqq \gamma$ and that $\| g \|_1 \leqq B =$ fixed constant (it is sufficient because of the finite memory that $\| g \|_2 \leqq$ fixed constant). We take as observation interval the interval $I = [0, T]$, and as the class S the set of all functions $x(t) \in L_2[0, T]$ which satisfy for a fixed σ, $0 < \sigma < 1$, and fixed $f_b > f_a > 0$,

$$\frac{2 \int_{f_a}^{f_b} | X(f) |^2 \, df}{\int_{-\infty}^{\infty} | X(f) |^2 \, df} = \frac{2 \int_{f_a}^{f_b} | X(f) |^2 \, df}{\| x \|^2} \geqq 1 - \sigma^2 \tag{11}$$

where

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{i2\pi f t} \, dt = \int_0^T x(t) e^{i2\pi f t} \, dt.$$

If σ is chosen too small the class S is empty, of course. We exhibit a uniform determination (ϵ, I), where $\epsilon = \epsilon(I)$, for the class of channels specified with respect to S. The idea is to transmit something like a "comb" of frequencies across the frequency band, $[f_a, f_b]$, of interest and measure the response to each. The transfer function cannot vary rapidly because of the finite memory, hence an approximation to the transfer function across the entire band can be obtained. This is the ordinary frequency-response method of testing a linear time-invariant system, just as the determination of example 1 is the ordinary impulse response-method.

First, by the assumption of finite memory, one has

$$| G(f + \xi) - G(f) | = \left| \int_0^\gamma [e^{i2\pi(f+\xi)} - e^{i2\pi f t}] g(t)\, dt \right| \leqq 2\pi\gamma \, | \xi | \, \| g \|_1 . \tag{12}$$

Now choose $\xi > 0$ arbitrarily; it is temporarily fixed. Let $\{f_i\}$ be a finite net of frequencies, $f_{i+1} - f_i = \xi_0$, $i = 1, \cdots, m$, such that $0 < f_1 - f_a \leqq \xi_0$, $0 < f_b - f_m \leqq \xi_0$. Let $x_i(t)$, $i = 1, \cdots, m$, be signals which vanish outside the interval $[0, T]$, each of which (for convenience) has the same L_2 norm. The ith signal is to be concentrated about the frequency f_i; that is, the number $\sigma_i > 0$ given by

$$\frac{2 \int_{f_i-\xi_0}^{f_i+\xi_0} | X_i(f)^2 | \, df}{\| x_i \|^2} \tag{13}$$

is to be small if possible. Put the transmitted signal $x(t) = \sum_{i=1}^m x_i(t)$ and put $p_i(t) = \alpha x_i(t)$ $(P_i(f) = \alpha X_i(f)$, where $P_i(f)$ is the Fourier transform of $p_i(t))$ where α is chosen so that

$$\int_{f_i-\xi_0}^{f_i+\xi_0} P_i(f) X_i(f)\, df = \alpha \int_{f_i-\xi_0}^{f_i+\xi_0} | X_i(f) |^2 \, df = \tfrac{1}{2}.$$

Then the quantity (p_i, Hx) is taken as an estimate of $G(f_i)$, and the step function

$$\tilde{G}(f) = (p_i, Hx), \qquad f_i - \xi_0 < f \leqq f_i + \xi_0, \qquad i = 1, \cdots, m$$

is an estimate of $G(f)$ over the interval $[f_i - \xi_0, f_m + \xi]$ which includes $[f_a, f_b]$. Outside this interval set $\tilde{G}(f) = 0$. Denote the union of the intervals $[f_i - \xi_0, f_i + \xi_0]$, and $[-f_i - \xi_0, -f_i + \xi_0]$ by Λ_i, then

$$G(f_i) - (p_i, Gx) = G(f_i) - \int_0^T p_i(t) \int_0^T g(t - \tau) x(\tau)\, d\tau\, dt$$
$$= G(f_i) - \int_{-\infty}^{\infty} \overline{P_i(f)}\, G(f) X(f)\, df$$

$$= \alpha \int_{\Lambda_i} | X_i(f) |^2 [G(f_i) - G(f)] \, df$$

$$+ \alpha \int_{\Lambda_i^*} | X_i(f) |^2 G(f) \, df$$

$$+ \alpha \int_{-\infty}^{\infty} \overline{X_i(f)} \sum_{k \neq i}^{m} G(f) X_k(f) \, df$$

where Λ_i^* is the set complementary to Λ_i. Hence,

$$| G(f_i) - (p_i , Gx) | \leqq 2\pi\gamma | \xi_0 | \| g \|_1 + \alpha \| g \|_1 \sigma_i \| x_i \|^2$$

$$+ \alpha \| g \|_1 \sum_{k \neq i}^{m} \left\{ \int_{\Lambda_i} | X_i(f) X_k(f) | \, df + \int_{\Lambda_i^*} | X_i(f) X_k(f) | \, df \right\}$$

where the fact that $| G(f) | \leqq \| g \|_1$ has been used. From the definition of the constants σ_i and α it follows that $\| x_i \|^2 = 1/\alpha(1-\sigma_i)$. Thus, by the Schwarz inequality,

$$| G(f_i) - (p_i , Gx) | \leqq \| g \|_1 \left\{ 2\pi | \xi_0 | \gamma + \frac{\sigma_i}{1 - \sigma_i} \right.$$

$$\left. + \sum_{k \neq i}^{m} \sqrt{\frac{\sigma_k}{1 - \sigma_k}} + \sum_{k \neq 1}^{m} \sqrt{\frac{\sigma_i}{(1 - \sigma_i)(1 - \sigma_k)}} \right\}$$

and hence from (12),

$$| G(f) - \tilde{G}(f) | \leqq | G(f) - G(f_i) | + | G(f_i) - (p_i , Gx) |$$

$$\leqq \| g \|_1 \left\{ 4\pi | \xi_0 | \gamma + \frac{\sigma_i}{1 - \sigma_i} + \sum_{k \neq i}^{m} \sqrt{\frac{\sigma_k}{1 - \sigma_k}} \right. \quad (14)$$

$$\left. + \sum_{k \neq i}^{m} \sqrt{\frac{\sigma_i}{(1 - \sigma_i)(1 - \sigma_k)}} \right\}.$$

Since $\| g \|_1 < B$ by assumption, the bound in (14) can be made arbitrarily small as $T \to \infty$. First, ξ_0 can be chosen to make the first term arbitrarily small. The choice of ξ_0 determines m. Then T can be chosen large enough to allow each of the σ_k to be arbitrarily close to 1. Estimates of possible values of the σ_k's for given ξ_0 and T are given by Pollak and Landau (1962). If we call the bound on $| G(f) - \tilde{G}(f) |$ given by (14), A, and denote the union of the intervals $[f_a , f_b]$, $[-f_b - f_a]$ by Λ then for $x \in S$,

$$\| Hx - \hat{H}x \|^2 = \int_{-\infty}^{\infty} | G(f) - \hat{G}(f) |^2 \cdot | X(f) |^2 \, df$$

$$\leqq A^2 \int_{\Lambda} | X(f) |^2 \, df + B^2 \int_{\Lambda^*} | X(f) |^2 \, df$$

$$\leqq (A^2 + B^2\sigma) \| x \|^2.$$

This kind of estimate is of interest when σ is very small, as it usually will be in examples from radio engineering. Of course, if the signals of interest tail off uniformly in energy away from a reference frequency, σ can be made arbitrarily small, for any T, by taking the band $[f_a, f_b]$ wide enough.

A class of channels is determinable only if the "degrees of freedom" are restricted in some way, or if there is the possibility that the measurement consists really of repeated measurements with the channel each time in the same state. One way to restrict the degrees of freedom is to require time invariance; another possible way would be to require all the channel operators to have the same principal axes, but this does not seem to be of practical interest; another way, which is a generalization of time invariance, is to require that all the channel operators evolve in time according to a known trend. We investigate now a notion of known trend. The idea is that if $h_0(t)$ is the response of a channel to an impulse occurring at time zero, the response to an impulse occurring at time s is to be given by a new function which is the result of a trend operator ψ_s operating on h_0.

For each s in some interval I (which may be infinite), let ψ_s be a bounded linear transformation with domain and range contained in $L_2[0, \gamma]$ which satisfies the conditions:

(i) Domain ψ_s = Domain $\psi_{s'}$ for all $s, s' \in I$.

(ii) For all $h \in$ Domain ψ_s and a. e. $t \in [0, \gamma]$

$$| (\psi_s h - \psi_{s'} h)(t) | \leqq \eta(s, s') \| h \|$$

where $0 \leqq \eta(s, s') = \eta(s', s) \rightarrow 0$ as $s' \rightarrow s$ and η is a continuous function of both variables for $s, s' \in I$.

(iii) For any $h, \tilde{h} \in$ Domain ψ_s

$$\| \psi_s h - \psi_s \tilde{h} \| \leqq \zeta(s) \| h - \tilde{h} \|$$

where $0 \leqq \zeta(s)$, $\zeta(s) \rightarrow 1$ as $s \rightarrow 0$ and $\zeta(s)$ is of integrable square on any finite interval $\subset I$. Then $\{\psi_s\}$ defines a class of channel kernels with *known trend* as follows: for any $h_0 \in$ Domain ψ_s put $h(t, 0) \equiv h_0(t)$ and then define

$$h(t, s) \equiv [\phi_s h_0](t - s), s \in I, t \in [s, s + \gamma].$$

Condition (i) is obviously necessary for the definition to make sense. Conditions (ii) and (iii) are more or less arbitrary continuity conditions chosen to guarantee that if two channel kernels are close together at one observation time they do not drift apart too rapidly, and to allow an easy characterization of determinable classes. In fact, if $h(u, 0) = g(u)$ belongs to $\mathcal{K}$, a uniformly determinable class $(0, T)$ of the type defined in Example 1, then $h(t, s) = [\phi_s h_0] (t - s)$ belongs to a uniformly determinable class $(0, T)$.

To prove this statement let $\{x_n(t)\}$ be a sequence of approximate δ-functions, as in 1, which vanish outside the interval $[0, \mu_n]$ and satisfy the condition that the product of the least upper bound of x_n, say B_n, and μ_n is bounded, $\mu_n B_n \leqq C$. Then μ_n has to approach zero, since x_n is an approximate δ-function, and one has

$$\left| \int_0^T h(t, s)x_n(s)\, ds - \int_0^T g(t - s)x_n(s)\, ds \right|$$

$$= \left| \int_0^{\mu_n} [(\phi_s h_0)(t - s) - h_0(t - s)]x_n(s)\, ds \right|$$

$$\leqq B_n \| h_0 \| \int_0^{\mu_n} \eta(s, 0)\, ds = B_n \mu_n \| h_0 \| \eta(\sigma_n, 0), \qquad 0 \leqq \sigma_n \leqq \mu_n,$$

by (ii). This approaches zero as $n \to \infty$, uniformly for h_0 in a determinable class of the type specified. Hence, in the space $L_2[0, T]$, $\| Hx_n - Gx_n \| = \| Hx_n - g{*}x_n \|$ is arbitrarily small for n sufficiently large, uniformly over $\mathcal{K}$. Here G is the convolution operator with kernel $g(u) \equiv h(u, 0)$, and H is the integral operator with kernel $h(t, s)$. Now suppose, given ϵ, n is large enough so that both

$$\| Hx_n - g{*}x_n \| < \epsilon$$

and

$$\| \phi_k - x_n{*}\phi_k \| < \epsilon/N, \qquad k = 1, \cdots, K,$$

where K and N satisfy the required inequalities in the proof of Example 1. Then consider the same determination as used there. One has

$$y_n = Hx_n = Gx_n + (H - G)x_n$$

so that

$$\sum_{k=1}^{K} (\phi_k, y_n)\phi_k = \sum_{k=1}^{K} (\phi_k, x_n*g)\phi_k + Y$$

where $\| Y \| \leqq \epsilon$ by the assumption above. Then

$$\left\| g - \sum_{k}^{K} (\phi_k, y_n)\phi_k \right\| = \left\| g - \sum_{k}^{K} (\phi_k, x_n*g)\phi_k - Y \right\|$$

$$\leqq \left\| g - \sum_{k}^{K} (\phi_k, x_n*g) \right\| + \epsilon,$$

which by the result in Example 1,

$$\leqq 5M\epsilon + \epsilon. \tag{15}$$

The determination $\sum_k^K (\phi_k, y_n)\phi_k$ yields an approximation to $h_0(t) = h(t, 0)$; we call this approximation $\hat{h}_0(t)$, i.e., $\hat{h}_0 = \sum_k^K (\phi_k, y_n)\phi_k$. The final determination of an approximate kernel is

$$\hat{h}(t, s) = (\psi_s \hat{h}_0)\ (t - s), \qquad 0 \leqq s < t \leqq T.$$

The square of the Hilbert-Schmidt norm of the error is then

$$\int_0^T \int_0^T | h(t, s) - [\psi_s \hat{h}_0](t - s) |^2\, dt\, ds$$

$$\leqq \int_0^T \int_s^{s+\gamma} | h(t, s) - [\psi_s \hat{h}_0](t - s) |^2\, dt\, ds$$

$$= \int_0^T \| \psi_s h_0 - \psi_s \hat{h}_0 \|^2\, ds \leqq \| h_0 - \hat{h}_0 \|^2 \int_0^T \zeta^2(s)\, ds$$

$$= \text{Const.} \cdot \| h_0 - \hat{h}_0 \|^2$$

and this, by (15), may be made uniformly small for $H \in \mathcal{H}$.

Example 4. This family of examples includes channel models for situations in which electromagnetic or sound radiation is scattered from a body of scatterers which is expanding or drifting. Given a constant $C > 0$, let $g(u)$ be a continuous function on the line vanishing outside $[0, \gamma]$ which satisfies the condition

$$| g(u) - g(u') | \leqq C \cdot \| g \| \cdot | u - u' |. \tag{16}$$

The set of such functions for which $\| g \|$ is less than a fixed bound, say

1, is a uniformly determinable class $(0, T)$ of time-invariant kernels. In fact, if one takes as orthonormal set on $[0, \gamma]$ the trigonometric functions

$$\sqrt{\frac{2}{\gamma}} \cos \frac{2\pi n t}{\gamma}, \qquad \sqrt{\frac{2}{\gamma}} \sin \frac{2\pi n t}{\gamma}, \qquad \frac{1}{\sqrt{\gamma}}, \qquad n = 1, 2, \cdots,$$

the nth Fourier coefficients of the class of g's satisfying (16) are dominated by $C' \| g \| (1/n)$ where $C' > 0$ is fixed (see, e.g., Titchmarsh (1939)). Hence the subclass with $\| g \| \leqq$ constant is a determinable class of the type of Example 1.

Now define $h(t, s)$ by

$$h(t, s) = [\psi_s g](t - s) = g(\alpha(s) + \beta(s)(t - s)) \tag{17}$$

where $\alpha(s)$, $\beta(s)$ are continuous in an interval I, $\alpha(0) = 0$, $\beta(0) = 1$ and $\beta(s)$ is bounded away from zero. Functions $h(t, s)$ as defined by Eq. (17) are channel kernels with known trend; condition (i) of the definition is obviously satisfied, conditions (ii) and (iii) may be verified easily (condition (ii) requires, of course, the Lipschitz condition on g). Furthermore, if one considers only those $g(u)$ satisfying the Lipschitz condition and $\| g \| \leqq$ constant, the class defined by Eq. (17) is uniformly determinable $(0, T)$.

The formal definitions of determinable class and determination have been introduced partly to indicate classes of channels for which an effective measurement is possible and partly to help keep straight the bookkeeping in an error analysis of such measurements. The idea being suggested here is that a channel measurement is feasible if the channel is known from prior information to belong to a specified uniformly determinable class or to be "near" such a determinable class, perhaps only in a statistical or average sense. Then a determination suitable to this class is used to estimate the actual channel operator. In this procedure errors may be caused for three reasons: 1) the presence of additive noise, 2) the fact that the channel being measured does not belong to the determinable class in question, but is only near to it, 3) the existence of residual error in measuring an element of the determinable class because of the finite nature of the determination, or because the class itself is too large. This third kind of error has already been discussed in the examples of determinable classes. We treat errors of the first two kinds essentially as perturbations on the measurement of channels belonging to a determinable class. Consider the second source of error. Suppose the observation interval is fixed, $I = [0, T]$; suppose the actual channel operator for this interval is H and that there is an operator H_0 in a bounded uniformly determinable class (ϵ, I) such that $\| H - H_0 \| < \eta$. Let $x(t)$ be the test signal, (p_k, w), $k = 1, \cdots, K$, the linear measurement and f the continuous function from K-dimensional space into L_2 which constitute the (ϵ, I)-determination. If there is no noise

$$w(t) = (Hx)(t) = (H_0x)(t) + (H - H_0)x(t) \tag{18}$$

and

$$(p_k, w) = (p_k, H_0x) + (p_k, (H - H_0)x).$$

Thus, putting $\epsilon_k = (p_k, (H - H_0)x)$, the determination yields

$$\begin{aligned} \hat{H} &= f((p_1, Hx), \cdots, (p_K, Hx)) \\ &= f((p_1, H_0x) + \epsilon_1, \cdots, (p_K, H_0x) + \epsilon_K) \end{aligned}$$

where

$$| \epsilon_k | \leqq \| H - H_0 \| \; \| x \| \; \| p_k \| \leqq \eta \| x \| \; \| p_k \|.$$

By hypothesis $\hat{H}_0$, defined to be $f((p_1, H_0x), \cdots, (p_k, H_0x))$, satisfies $\| \hat{H}_0 - H_0 \| < \epsilon$, hence

$$\begin{aligned} \| \hat{H} - H \| &\leqq \| \hat{H} - \hat{H}_0 \| + \| \hat{H}_0 - H_0 \| + \| H_0 - H \| \\ &\leqq \| f((p_1, H_0x) + \epsilon_1, \cdots, (p_K, H_0x) + \epsilon_K) \\ &\qquad - f((p_1, H_0x), \cdots, (p_K, H_0x)) \| + \epsilon + \eta. \end{aligned}$$

Now, since f is a known operator-valued function which is uniformly continuous on any closed bounded set in K-dimensional Euclidean space, given any $\delta > 0$ there is an $\eta_0 > 0$ such that for η small enough so that $\eta \| x \| \; \| p_k \| \leqq \eta_0$, the above inequality reduces to

$$\| \hat{H} - H \| \leqq \delta + \epsilon + \eta.$$

We have actually proved that the class of all H which are within a distance η of a bounded uniformly determinable class (ϵ, I), are themselves a uniformly determinable class (ϵ', I), where $\epsilon' \to \epsilon$ as $\eta \to 0$.

If there is noise present, or if the channel is known to be in or near a determinable class only in a statistical sense, we can no longer establish sure error bounds, but can only make probabilistic statements about error bounds. To illustrate this point, let us consider the case where the channel is characterized stochastically and we know only that for some $\alpha > 0$, H satisfies $E \| H - H_0 \|^2 < \alpha$, for some H_0 in a prescribed bounded uniformly determinable class (ϵ, I). Further, let us suppose there is additive noise present and we know that for some $\beta > 0$, $E \| n \|^2 < \beta$. One has

$$(p_k, w) = (p_k, H_0 x) + (p_k, (H - H_0)x) + (p_k, n).$$

Let η_1, η_2 be arbitrary positive numbers. Then

$$P\{\| H - H_0 \| \leqq \eta_1\} \geqq 1 - \frac{\alpha^2}{\eta_1^2}$$

$$P\{\| n \| \leqq \eta_2\} \geqq 1 - \frac{\beta^2}{\eta_2^2},$$

and if these events are independent, one can say that with probability exceeding $[1 - (\alpha^2/\eta_1^2)]\,[1 - (\beta^2/\eta_2^2)]$,

$$\begin{aligned} \| \hat{H} - H \| \leqq \| f((p_1, H_0 x) + \epsilon_1, \cdots, (p_K, H_0 x) + \epsilon_K) \\ - f((p_1, H_0 x), \cdots, (p_K, H_0 x)) \| + \epsilon + \eta_1 \end{aligned} \tag{19}$$

where $| \epsilon_k | \leqq \eta_1 \| x \| \; \| p_k \| + \eta_2 \| p_k \|$. Again, the right side of the inequality (19) approaches ϵ as $\eta_1 \; \eta_2 \to 0$, but, of course, the error bound is valid with probability nearly one only if α, β are small. The condition that the determinable class be bounded can be dropped by replacing the first term of the inequality above by

$$\| f((p_1, w), \cdots, (p_k, w)) - f((p_1, w) - \epsilon_1, \cdots, (p_K, w) - \epsilon_K) \|$$

but then the bound is no longer uniform, and the (p_i, w) must be known before the bound can be determined.

It is worth remarking that if preliminary smoothing filtering is done to minimize the relative noise intensity, the smoothing filters in cascade with the original channel define a new channel to be determined as above.

IV. MEASUREMENT AND USE OF SLOWLY-VARYING CHANNELS

We consider now channels which are varying with time in an unknown fashion, but at a sufficiently slow rate to permit approximation over a useful interval by time-invariant channels, or, more precisely, by integral operators with time-invariant kernels. If a channel is alternately proved and used as a medium for communication, there are errors introduced, first in the channel measurement, and second in the extrapolation of the measured channel characteristics into the near future. It is proposed to treat this situation in a way which is partly statistical and partly deterministic, and which uses the ideas of the preceding section.

We suppose that any transmitted signal $x(t)$, $\tau_1 \leqq t \leqq \tau_2$, results in a received signal of the form

$$w(t) = \int_{\tau_1}^{\tau_2} h(t, s)x(s)\, ds + n(t), \qquad \tau_1 \leqq t \leqq \tau_2$$

where $n(t)$ is noise (to be specified in more detail later) and $h(t, s)$, defined for $-\infty < t, s < \infty$, is an ot-stationary stochastic kernel with mean zero which characterizes the channel. That is, we suppose (presumably from some knowledge of the physics of the channel, and preliminary statistical tests) that it is reasonable to model the channel as an ot-stationary stochastic channel and that the channel autocorrelation function,

$$R(\tau; u, v) = Ek(t + \tau, u)k(t, v)$$

is known, at least to a rough approximation. If the autocorrelation function $R(\tau; u, v)$ has certain properties, the channel will be slowly-varying on the average, so it makes sense to approximate the sample functions of $k(t, u)$ by time-invariant kernels for t-intervals that are not too long. Measurement procedures can be given, based on the results of the preceding section, which will yield such a time-invariant approximation. This approximation to the channel can then be used in the signal processing when the channel is used as a communication medium.

We refer to the Appendix for proofs and elaboration of the following facts about approximation by time-invariant kernels and the connection with ot-stationarity:

(i) Let $k(t, u)$ be an admissible kernel. The time-invariant kernel $g(u)$ which most closely approximates $k(t, u)$ in HS norm on the interval $[0, T]$, i.e., which minimizes

$$\int_0^T \int_0^t | k(t, u) - g(u) |^2 \, dt\, du,$$

is given by

$$g(u) = \frac{1}{T - u} \int_u^T k(\tau, u)\, d\tau. \tag{A.1}$$

(ii) If $k(t, u)$ is an admissible ot-stationary stochastic kernel, the stochastic process $g(u)$ which best approximates (t, u) in the sense of minimizing

$$E \int_0^T \int_0^t |\, k(t, u) - g(u) \,|^2 \, dt \, du \tag{20}$$

is still given by Eq. (A.1), and the error, that is the value of (20), is given by

$$\int_0^T \int_0^T \int_0^{\min(t,\tau)} \frac{\rho(\tau - t, u)}{T - u} \, du \, d\tau \, dt \tag{A.7}$$

where $\rho(t, u) \equiv R(0; u, u) - R(t; u, u)$.

Example 1 (*cont.*). Consider a stochastic channel kernel function $k(t, u)$ which is ot-stationary, and which has the property that with probability one for a.e. t, $k(t, \cdot)$ satisfies the condition of uniform domination of Fourier coefficients with respect to some cons $\{\phi_k\}$ required in Example 1. Then the best time-invariant approximation of $k(t, u)$, given by Eq. (A.1), satisfies this condition, as does also the simple approximation given by simply fixing t in $k(t, u)$ (for a.e. t). Thus a uniformly determinable class of the type of Example 1 is appropriate, and $\| \hat{H} - H \|$ satisfies the inequality (20) with the probability stated. The number α, which is a bound on $E \| H - H_0 \|^2$, can be taken from Eq. (A.7) or (A.9), where the former gives the best possible (i.e., the smallest) value. The number β, which is a measure of noise intensity, must be a datum of the problem. Since in this example the determination is simply a partial sum of a Fourier series, one has for the $\hat{H} - \hat{H}_0$ contribution to the error, where $\hat{g}(u)$ and $\hat{g}_0(u)$ are the time-invariant kernels for $\hat{H}$ and $\hat{H}_0$ respectively,

$$\begin{aligned} \hat{g} - \hat{g}_0 &= \sum^K (\phi_k, w)\phi_k - \sum^K (\phi_k, H_0 x)\phi_k \\ &= \sum^K [(\phi_k, (H - H_0)x) + (\phi_k, n)]\phi_k, \end{aligned}$$

whence

$$\| \hat{g} - \hat{g}_0 \| \leqq \| H - H_0 \| \; \| x \| + \| n \|,$$

and

$$[\![\hat{H} - \hat{H}_0]\!] \leqq \sqrt{T} \, [\, \| H - H_0 \| \; \| x \| + \| n \| \,].$$

Thus, from (20), one can say that with probability exceeding

$$[1 - (\alpha^2/\eta_1^2)]\,[1 - (\beta^2/\zeta^2)]$$

$$\| \hat{H} - H \| \leqq \eta_1 \sqrt{T}\, \| x \| + \sqrt{T}\zeta + \epsilon + \eta_1 \tag{21}$$

where ϵ is the residual error in the determination, $\eta_1 > \alpha$, $\eta_2 > \beta$ are arbitrary and α, β are given as above. One should recall that a lower bound for $\| x \|$ in this inequality is fixed by the choice of ϵ; once a residual error ϵ is established a sufficiently good approximate δ-function $x(t)$ is required. Since the L_1-norm of $x(t)$ must be held at one (because of the normalization of the ϕ_k) the L_2-norm of x must exceed some lower bound. Note that any H_0 in the determinable class may be used in obtaining the estimate (21) if α is chosen appropriately. The best estimate of this kind is obtained with minimum α, and this is achieved in this example by using for H_0 the best approximation as given by Eq. (A.1), which implies, as already stated, that α can be given by Eq. (A.7).

In order to bound the error which occurs in using the measured value of the channel kernel function in the immediate future one need modify the inequality (21) only slightly. Let us suppose the channel has finite memory γ, that it is to be measured during the interval $[0, a]$ and used during the interval $[b, b + T]$, when $0 < a < b < b + T$. The measurement is to be accomplished by reference to the same determinable class as specified in the preceeding paragraph.

Let H be the actual channel operator during the measurement interval and H_c the actual channel operator during the use interval. H_0 is to be an operator for the use interval whose kernel is a time-invariant approximation to that of H belonging to the same determinable class as above; $\hat{H}_0$ is the estimate of H_0 which would be yielded by the determination if H_0 were the actual channel operator and if there were no noise, and $\hat{H}$ is the estimated operator. H_0' is a time-invariant operator for the measurement interval with the same kernel as H_0. Suppose,

$$E \| H - H_0' \|^2 < \alpha_1^2$$

$$E \| H_c - H_0 \|^2 < \alpha_2^2$$

and η_1, η_2, are arbitrary positive numbers. Then,

$$P\{\| H - H_0 \| \leqq \eta_1 , \| H_c - H_0 \| \leqq \eta_2\} \geqq 1 - \frac{\alpha_1^2}{\eta_1^2} - \frac{\alpha_2^2}{\eta_2^2} ,$$

and since

$$\hat{H} - H_c = (\hat{H} - \hat{H}_0) + (\hat{H}_0 - H_0) + (H_0 - H_c)$$

one can say by an argument paralleling the previous one that, with probability exceeding

$$\left(1 - \frac{\alpha_1^2}{\eta_1^2} - \frac{\alpha_2^2}{\eta_2^2}\right)\left(1 - \frac{\beta^2}{\zeta^2}\right), \qquad (22)$$
$$\| H - H_c \| \leqq \eta_1 \sqrt{T} \| x \| + \zeta\sqrt{T} + \epsilon + \eta_2 .$$

In the previous paragraph it was pointed out that the inequality (21) was derivable with suitable constants no matter what H_0 was used for comparison, but that the best result was obtained if $E \| H - H_0 \|^2$ was a minimum. Again, (22) follows with suitable constants for any H_0 in the determinable class, but it is no longer clear what is the best choice of H_0 in deriving the inequality since $\| \hat{H} - \hat{H}_0 \|$ and $\| H_0 - H_c \|$ can have quite complicated behavior relative to each other as H_0 is varied, depending on the actual autocorrelation function. If, for example, we take for H_0 the integral operator on $L_2[b, b + T]$ with kernel $k(c, u)$ where c is a constant, $0 \leqq c \leqq b + T$, then by (A.10) α_1 and α_2 can be taken

$$\alpha_1^2 = 2 \int_0^a [R(0) - R(t - c)]\, dt$$

$$\alpha_2^2 = 2 \int_b^{b+T} [R(0) - R(t - c + b)]\, dt.$$

There is a hidden constraint on the measurement interval $[0, a]$ which is implicit in these inequalities. Suppose $T > \gamma$. Then since H_0 is an operator on an interval of length T, determining $\hat{H}_0$ so that $\| \hat{H}_0 - H_0 \| \leqq \epsilon$ as required necessitates a measurement interval of length nearly γ, and its length must be $\geqq \gamma$ in the limit as $\epsilon \to 0$. Thus, practically one can say that $a > \gamma$. For a to be greater than is necessary to make the determination weakens the error inequality (22), however, by increasing α_1. Thus, the interpretation of (22) agrees with the common sense idea that one can apply the test signal, take measurements until the channel stops ringing, then use the channel until it has drifted far enough to cause an unacceptable error.

The successive measurements of the channel are, of course, available for improving an estimate of $R(t; u, v)$, but that aspect of the problem will not be discussed here.

Example 3. (*cont.*). This will just be indicated. Let a stochastic kernel function which is ot-stationary have finite memory $\leqq \gamma$. Then again since averaging on t preserves the finite memory property, this channel can be referred to a determinable class of the type discussed in Example 3. Then the probabilistic inequality of (19) is valid, where the determination referred to is that of Example 3. In this case, of course, H and $\hat{H}$ are restricted to the subset of nearly-band-limited signals introduced in Example 3. The error which occurs in using the estimated channel operator at a future time is subject to bounds established in the same way as in the example above.

V. AN APPLICATION TO SIGNAL DETECTION AND MEASUREMENT

The material of the preceding sections is intended to describe classes of channels which can be measured approximately, and to provide estimates for how much the actual channel operation may differ from what the receiver thinks it is. Results of this kind can be used to show when in signal-detection problems certain standard statistical data-processing procedures may be used, if information about the channel is continuously updated, and how much loss in performance may be incurred because of the time-varying nature of the channel. We illustrate this application in this section with a known example (Grenander, 1949).

Let the received signal be

$$w(t) = y(t; \alpha) + n(t), \qquad \tau_1 \leqq t \leqq \tau_2 \tag{1}$$

where now we fix the noise $n(t)$ to be a Gaussian process continuous in mean-square and with mean zero, and $y(\cdot; \alpha)$ to be a known real-valued function $\epsilon L_2[\tau_1, \tau_2]$ for each α in a parameter set A, where A is either a finite set or a compact subset of R_k. Let $R(t, s) = En(t)n(s)$, and let

$$\int_{\tau_1}^{\tau_2} R(t, s)\phi_n(s)\, ds = \lambda_n \phi_n(t), \qquad \tau_1 \leqq t \leqq \tau_2 .$$

The λ_n are nonnegative; we shall assume the $\{\phi_n(t)\}$ are taken to be orthonormal and real and that the integral operator in question has zero null space, so that $\{\phi_n\}$ is a complete set. We define

$$w_k = \int_{\tau_1}^{\tau_2} w(t)\phi_k(t)\, dt$$

$$y_k(\alpha) = \int_{T_1}^{T_2} y(t; \alpha)\phi_k(t)\, dt \tag{23}$$

$$f(w; \alpha) = \sum^{\infty} \frac{w_k s_k(\alpha)}{\lambda_k}.$$

The w_k are jointly Gaussian random variables. If for each $\alpha \in A$. $\sum y_k^2(\alpha)/\lambda_k < \infty$, then the series defining $f(w; \alpha)$ converges with probability one and also in mean square with respect to the measure induced by any $\alpha \in A$. Also

$$E_\alpha f(w; \alpha_0) = \sum^{\infty} \frac{y_n(\alpha)y_n(\alpha_0)}{\lambda_n} \tag{24}$$

$$\operatorname{var}_\alpha f(w; \alpha_0) = \sum^{\infty} \frac{y_n^2(\alpha_0)}{\lambda_n} \tag{25}$$

and $f(w; \alpha)$ is Gaussian. The subscript α refers to the measure induced by the parameter α. Then the logarithm of the "likelihood ratio," i.e., the logarithm of the Radon-Nikodym derivative of the two probability measures induced on the sample space of the $w(t)$ by the parameters α_1 and α_0 is given by

$$\lim_{N\to\infty} \log \frac{p(w_1, \cdots, w_N; \alpha_0)}{p(w_1, \cdots, w_N; \alpha_1)} = f(w; \alpha_1) - f(w; \alpha_0) + C(\alpha_1, \alpha_0)$$

where $C(\alpha_1, \alpha_0)$ depends on $s(t; \alpha_0)$, $s(t; \alpha_1)$ but not on $w(t)$. Thus any inference procedure (i.e., hypothesis test or point estimation) based on likelihood ratios is determined by the test functionals $f(w; \alpha)$. The behavior of any such inference procedure depends on the distributions of the $f(w; \alpha)$; and since these are all jointly Gaussian, on the first and second moments of the $f(w; \alpha)$. Thus, for the class of sure-signal-in-noise problems indicated, and from an applicational point of view this is a wide class, one can investigate the effect of unknown perturbations on the prior data of the problem entirely by first and second moment calculations of the $f(w; \alpha)$. Such a problem is said to be stable (5) if a small change in the noise covariance (in the sense of L_2-norm) necessarily causes only a small change in the distribution functions of the $f(w; \alpha)$. A necessary and sufficient condition for stability is that

$$\sum^{\infty} \frac{y_n^2(\alpha)}{\lambda_n^2} < \infty. \tag{26}$$

If this condition (26) holds then it also follows immediately that for any α the mean value of the test functional $f(w;\alpha_0)$ varies continuously with perturbations of the signal $y(t;\alpha_0)$, where again the L_2-norm is used to measure the perturbations. In fact, let $y'(t;\alpha) = y(t;\alpha) + e(t;\alpha)$, where $e(t;\alpha)$ is to be regarded as a perturbation, be the actual received signal so that

$$w'(t) = y'(t;\alpha) + n(t).$$

The test functionals $f(\cdot;\alpha_0)$ are unchanged because they represent fixed data processing procedures. However, their mean values are changed,

$$E_\alpha f(w', \alpha^0) = E_\alpha \sum^{\infty} \frac{w_k' y_k(\alpha_0)}{\lambda_k} \tag{27}$$

$$= \sum^{\infty} \frac{y_k'(\alpha) y_k(\alpha_0)}{\lambda_k} \tag{28}$$

where

$$y_k'(\alpha) = \int_{\tau_1}^{\tau_2} e(t;\alpha)\phi_k(t)\,dt + y_k(\alpha)$$

$$= e_k(\alpha) + y_k(\alpha).$$

Thus

$$E_\alpha f(w';\alpha^0) = E_\alpha f(w;\alpha^0) + \sum^{\infty} \frac{e_k(\alpha) y_k(\alpha_0)}{\lambda_k},$$

and the absolute value of the change in the mean is

$$\leqq \| e(\alpha) \| \sum^{\infty} (y_k^2\, \alpha_0)/\lambda_k^2)^{1/2}.$$

The variances of the $f(\cdot;\alpha)$ are unaffected by changes in actual received signal.

The simplest example in which to carry through the effect of perturbations on the final inference is a pure detection problem, but even though simple it illustrates the situation adequately. Let $\alpha = 0$ or 1, take $y(t;0) = s(t)$, $\tau_1 \leqq t \leqq \tau_2$, a known function, and take $y(t;1) = 0$. Then a likelihood test for the presence of the signal $s(t)$ is to compare $f(w;0)$ with a fixed threshhold for η, and decide the signal is present if $f(w;0) > \eta$. One has then that the probability of correctly deciding that the signal is present is

$$P\{f(w; 0) > \eta \mid \alpha = 0\} = \frac{1}{\sqrt{2\pi}} \int_{\eta/\sqrt{b}-\sqrt{b}}^{\infty} e^{-u^2/2}\, du$$

where $b = \sum^{\infty} y_k^{2(0)}/\lambda_k$ is the signal-to-noise ratio. Now suppose that the actual signal is $s(t) + e(t)$, and put

$$b' = E_0 f(w; 0) = + \sum \frac{e_k y_k(0)}{\lambda_k}.$$

Then

$$P\{f(w; 0) > \eta \mid \alpha = 0\} = \frac{1}{\sqrt{2\pi}} \int_{(\eta-b')/\sqrt{b}}^{\infty} e^{-u^2/2}\, du.$$

The change in the lower limit of the error function integral is

$$-\sum \frac{e_k y_k(0)}{\lambda_k} \Big/ \sum \left(\frac{y_k^2(0)}{\lambda_k}\right)^{1/2} \tag{29}$$

and the absolute value of this change is less than or equal to

$$\| e \| \left[\frac{\sum^{\infty} y_k^2(0)/\lambda_k^2}{\sum^{\infty} y_k^2(0)/\lambda_k}\right]^{1/2}. \tag{30}$$

Of course the effect of this perturbation on the probability of detection depends where on the tail of the Gaussian distribution $(\eta - b)/\sqrt{b}$ is located, and as the signal-to-noise ratio b becomes larger the effect is less.

Now for the time-varying linear channel

$$y(t; \alpha) = \int_{\tau_1}^{\tau_2} h(t, s) x(s; \alpha)\, ds, \qquad \tau_1 \leqq t \leqq \tau_2 \tag{31}$$

and if $h(t, s)$ is known (it is assumed $x(s; \alpha)$ is known) one has the necessary prior data on the signal for a sure-signal-in-noise problem. The application of channel measurement techniques is obvious. One uses the estimated channel kernel $\hat{h}(t, s)$ (or $\hat{k}(t, u)$) to yield a *nominal* received signal for each α:

$$y_0(t; \alpha) = \int_{\tau_1}^{\tau_2} \hat{h}(t, s) x(s; \alpha)\, ds, \qquad \tau_1 \leqq t \leqq \tau_2.$$

The data processing is based on $y_0(t; \alpha)$. The actual received signal function $y(t; \alpha)$ is given by Eq. (31), and $e(t; \alpha) = y(t; \alpha) - y_0(t; \alpha)$, is the difference between it and the nominal signal. Then,

$$\| e(\cdot\, ; \alpha)\| \leqq \| H - \hat{H} \| \, \| x(\cdot\, ; \alpha \|.$$

From results of the type of those obtained in Section IV, one can say that for certain numbers ϵ, $A(\alpha) > 0$, $\| e(\cdot\, ; \alpha)\| \leqq A(\alpha)$ with probability $\geqq 1 - \epsilon$, and hence with probability $\geqq 1 - \epsilon$ the mean-value of the test functional $f(w; \alpha)$ is changed by less than

$$A(\alpha) \sum^{\infty} \left(\frac{y_{k0}^2(\alpha)}{\lambda_k^2} \right)^{1/2} \tag{32}$$

from its nominal value, where $y_{k0}(\alpha)$ is the Fourier coefficient with respect to ϕ_k of the nominal signal $y_0(t; \alpha)$. It should be mentioned that the factor

$$\sum^{\infty} \frac{y_{k0}^2(\alpha)}{\lambda_k^2} \tag{33}$$

can also be written as $\| z(\cdot\, ; \alpha)\|^2$ where $z(t; \alpha)$ is the solution of

$$\int_{\tau_1}^{\tau_2} R(t, s) z(s; \alpha)\, ds = y_0(t; \alpha), \tau \qquad \tau_r \leqq t \leqq t_2 .$$

This equation has a solution in $L_2[\tau_1, \tau_2]$ if the series in (26) converges, and it is unique by the assumption that the integral operator has zero null space.

APPENDIX: THE APPROXIMATION OF OBSERVATION-TIME-STATIONARY KERNELS BY TIME-INVARIANT KERNELS

The best mean-square approximation to an arbitrary realizable kernel by a time-invariant kernel is obtained by averaging the original kernel over the observation time. More precisely, one can state the following:

Lemma. If $k(t, u) \in L_2[0, T] \times [0, T]$ and $k(t, u) = 0$ for all $u > t$, then

$$g(u) = \frac{1}{T - u} \int_u^T k(\tau, u)\, d\tau \tag{A.1}$$

is defined for a.e. u, $0 \leqq u \leqq T$,

$$\int_0^T \int_0^t g^2(u)\, du < \infty, \tag{A.2}$$

and amongst all functions $g'(u)$ satisfying (A.2), *g provides a minimum for the expression*

$$\int_0^T \int_0^t | k(t, u) - g'(u) |^2 \, du \, dt \equiv [\![k - g]\!]^2. \tag{A.3}$$

In other words, $g(u)$ is the kernel for the realizable, time-invariant *integral operator which most closely approximates the channel operator in Hilbert-Schmidt norm.*

Proof: Since $k(t, u)$ is of integrable square on $[0, T] \times [0, T]$ it is integrable on $[0, T] \times [0, T]$ and hence $g(u)$ is defined for a.e. u. We calculate a bound for the expression (A.3) with $g(u)$ used for $g'(u)$:

$$\begin{aligned}
[\![k - g]\!]^2 &= \int_0^T \int_0^t \left| \frac{1}{T - u} \int_u^T [k(t, u) - k(\tau, u)] \, d\tau \right|^2 du \, dt \\
&= \int_0^T \int_u^T \frac{1}{(T - u)^2} \int_u^T [k(t, u) - k(\tau, u)] \, dt \\
&\quad \cdot \int_0^T [k(t, u) - k(\tau', u)] \, d\tau' \, dt \, du \\
&= \int_0^T \frac{1}{(T - u)^2} \int_u^T \int_u^T \int_u^T \{k^2(t, u) \; k(t, u)k(\tau', u) \\
&\quad - k(t, u)k(\tau, u) + k(\tau, u)k(\tau', u)\} \, d\tau \, d\tau' \, dt \, du \qquad \text{(A.4)} \\
&= \int_0^T \left\{ \int_u^T k^2(t, u) \, dt - \frac{1}{T - u} \right. \\
&\quad \left. \cdot \int_u^T \int_u^T k(t, u)k(\tau, u) \, dt \, d\tau \right\} du \\
&= [\![k]\!]^2 - \int_0^T \left[\int_u^T \frac{k(t, u)}{\sqrt{T - u}} \, dt \right]^2 du.
\end{aligned}$$

Thus $[\![g]\!] \leqq 2 [\![k]\!]$ and (A.2) is satisfied. Furthermore, it follows from the Schwarz inequality applied to the last term in Eq. (A.4) that $\| k - g \| = 0$ if and only if $k(t, u)$ does not depend on t. Now $g' = g$ will minimize $[\![k - g]\!]$ if

$$\int_0^T \int_0^t | k(t, u) - g(u) - \eta(u) |^2 du \, dt \geqq \int_0^T \int_0^T | k(t, u) - g(u) |^2 \, du \, dt$$

for any $\eta(u) \in L_2[0, T]$. This condition reduces to the requirement that

$$\int_0^T \int_0^t \eta(u)[k(t, u) - g(u)] \, du \, dt = 0$$

or

$$\int_0^T \int_0^t \frac{\eta(u)}{T-u} \int_u^T [k(t, u) - k(\tau, u)] \, d\tau \, du \, dt = 0$$

for any $\eta(u) \in L_2[0, T]$. This integral can be rewritten

$$\int_0^T \int_0^T \int_0^{\min(t,\tau)} \frac{\eta(u)}{T-u} [k(t, u) - k(\tau, u)] \, du \, d\tau \, dt. \tag{A.5}$$

The bracketed expression in (A.5) is an antisymmetric function of t, τ; hence, since the double integral in (A.5) is over the square $0 \leqq t \leqq T$, $0 \leqq \tau \leqq T$, it vanishes for all $\eta(u)$ as required.

If $k(t, u)$ is a stochastic kernel satisfying the condition (A.3) then it follows immediately that $g(u)$ as given by Eq. (A.1) is defined, except for a set of realizations of $k(t, u)$ of probability zero, for a.e. u, and minimizes $E \| k - g \|^2$ within the class of all $g(u)$ satisfying

$$E \int_0^T \int_0^t g^2(u) \, du < \infty.$$

The mean-square error of approximation of an ot-stationary kernel by the best time invariant one as given by Eq. (A.1), i.e., the expected value of the HS norm of the difference, is given for an arbitrary interval $[a, a + T]$ by

$$\begin{aligned}
E \int_a^{a+T} \int_0^{t-a} &\left| k(t, u) - \frac{1}{T-u} \int_{a+u}^{a+T} k(\tau, u) \, d\tau \right|^2 du \, dt \\
&= E \int_0^T \frac{1}{(T-u)^2} \int_{a+u}^{a+T} \left| \int_{a+u}^{a+T} [k(t, u) - k(\tau, u)] \, d\tau \right|^2 dt \, du \\
&= \int_0^T \frac{1}{(T-u)^2} \int_{a+u}^{a+T} \int_{a+u}^{a+T} \int_{a+u}^{a+T} \{R(0; u, u) - R(t - \tau'; u, u) \\
&\qquad - R(\tau - t; u, u) + R(\tau - \tau'; u, u)\} \, d\tau \, d\tau' \, dt \, du \\
&= \int_0^T (T - u) R(0; u, u) \, du \\
&\qquad - 2 \int_0^T \frac{1}{T-u} \int_u^T \int_u^T R(\tau - t; u, u) \, d\tau \, dt \, du \\
&\qquad + \int_0^T \frac{1}{T-u} \int_u^T \int_u^T R(\tau - \tau'; u, u) \, d\tau \, d\tau' \, du \\
&= \int_0^T (T - u) R(0; u, u) \, du \\
&\qquad - \int_0^T \int_u^T \int_u^T \frac{1}{T-u} \{R(\tau - t; u, u)\} \, d\tau \, dt \, du.
\end{aligned} \tag{A.6}$$

One notices in Eq. (A.6) that the mean-square error of the time-invariant approximation does not depend on the translation parameter a; this is true because of the ot-stationarity and would be expected. The right side of Eq. (A.6) may be rewritten as

$$\int_0^T \int_u^T \int_u^T \frac{\rho(\tau - t, u)}{T - u} \, d\tau \, dt \, du \tag{A.7}$$

where $\rho(t, u) \equiv R(0; u, u) - (R(t; u, u)$ is a statistical measure of how rapidly the channel is varying.

The rather awkward expression (A.7) can sometimes be replaced by a somewhat crude but very simple upper bound as follows. Let $\delta = \min[T, \gamma]$, γ the upper bound on the memory of $k(t, u)$. Then

$$\begin{aligned} E \int_0^T \int_0^t \left| k(t, u) - \frac{1}{T - u} \int_u^T k(\tau, u) \, d\tau \right|^2 du \, dt \\ \leqq E \int_0^T \int_0^\delta \left| k(t, u) - \frac{1}{T} \int_0^T k(\tau, u) \, d\tau \right|^2 du \, dt, \end{aligned} \tag{A.8}$$

since the integrand is positive and since $(1/T) \int_0^T k(\tau, u) \, d\tau$ is a time-invariant kernel yielding no better approximation than the optimum one. The right side of (A.8) can be easily evaluated to give

$$TR(0) - 2 \int_0^T R(\eta) \, d\eta + \frac{2}{T} \int_0^T \eta R(\eta) \, d\eta \tag{A.9}$$

where

$$R(\eta) \equiv \int_0^\delta R(t; u, u) \, du.$$

Now if the condition $R(0) - R(\eta) \leqq \epsilon$ for $|\eta| \leqq T$ is satisfied, the error bound (A.9) is less than or equal to $2 \epsilon T$.

In general, the mean-square error (in the sense we have been using that term here) in approximating a stochastic kernel over a finite interval by any linear transformation of the same kernel can obviously be expressed in terms of its autocorrelation function. One other simple example of this, which is used here, is

$$\begin{aligned} \int_0^T \int_0^t | k(t, u) - k(b, u) |^2 \, du \, dt \\ = 2 \int_0^T \int_0^t [R(0; u, u) - R(t - b; u, u)] \, du \, dt \\ \leqq 2 \int_0^T [R(0) - R(t - b)] \, dt \end{aligned} \tag{A.10}$$

where $R(\eta)$ is defined as before.

RECEIVED: July 16, 1964

REFERENCES

BELLO, P. A. (1962), "Characterization of Randomly Time-Variant Linear Channels," ADCOM Research Rept. #2.

GRENANDER, U. (1950), Stochastic processes and statistical inference. *Ark. Mat.* **1**, 195–277.

KAILATH, T. (1963), Time-variant communication channels. *IEEE Trans. Inform. Theory* **IT-9**, 233–237.

KAILATH, T. (1959), "Sampling Models for Linear Time-Variant Filters," M.I.T. Research Laboratory of Electronics, Tech. Rept. #352.

KOLMOGOROV AND TIHOMOROV, (1959), Entropy and capacity of sets in functional spaces. *Transl. Am. Math. Soc.* **31**.

LANDAU, H. J., AND POLLAK, H. O. (1962), Prolate spheroidal wave functions, Fourier analysis and uncertainty—III. *Bell System Tech. J.* **41**, 1295–1336.

PRICE, R., AND GREEN, JR., P. E. "Signal Processing in Radar Astronomy—Communication via Fluctuating Multipath Media". M.I.T., Lincoln Laboratory Tech. Rep. 234, October, 1960.

ROOT, W. L. (1963), Stability in signal detection problems. *Proc. Symp. Appl. Math.* **16**, New York.

TITCHMARSH, E. C. (1932), "The Theory of Functions." Oxford Univ. Press.