cess of Newtonian mechanics, it seemed tempting to brush away the complex and eroded conceptual system of scholastic philosophy. The proposition of temporarily restricting oneself to the conceptual framework of the new mechanics seems entirely sound, even in retrospect. However, the contention that this conceptual framework would be satisfactory at all times was unwarranted, and turned out to be actually incorrect.

During the mechanistic era it became customary to dismiss types of questions that did not fit into mechanistic systems as unscientific.

As the crisis of classical physics revealed the limitation of the mechanistic conceptual scheme, the first inference was that the range of *legitimate scientific questions is even further limited,* since not even the mechanistic questions are admissible.

The pluralistic character of the present approach brings two new elements into this picture. In the first place, each deductive system implies a characteristic set of precise questions. The number of interesting questions that become "meaningful" is particularly extensive in thermodynamics and quantum mechanics.

It is often stated that the concept of object breaks down in quantum mechanics. Actually, however, the opposite is true. As we have seen in Sec. IV, for the first time in QM we are in a position to give a formal representation of an object with many subtle ramifications and we can now solve the related philosophical puzzles that have been unresolved since their discovery by the Eleatic philosophers. It seems that the new object concept is flexible enough to include living organisms that are entirely outside the mechanistic scheme.

The extension of meaningful conceptual problems in the present context proceeds in still another dimension. Not only do we have the concepts within each deductive system, but the deductive systems themselves are conceptual entities of distinct individual characteristics related to each other in quite specific fashion. These entities are of a logical type that is markedly different from that of the primitive concepts within the deductive systems.

REVIEWS OF MODERN PHYSICS VOLUME 35, NUMBER 1 JANUARY 1963

Linear Stochastic Operators*

G. ADOMIAN
Hughes Aircraft Company, Culver City, California

INTRODUCTION

A GENERAL class of linear stochastic operators arises in problems involving propagation in a random medium and in the processing (or observation) of an input quantity (which may itself be a random process) by a linear system whose parameters may be stochastic.

For example, in Mie scattering from aerosols in the atmosphere, we might wish to give a probabilistic treatment to the normally assumed uniform distributions of particle size and spacing. We might investigate ionospheric effects on satellite communications due to ionization variations brought about by turbulence, upper atmospheric winds, or magnetohydrodynamic interactions. We may want to inquire into the ultimate accuracy and sensitivity in various measurement processes basically subject to random perturbations.

Applications of stochastic operators to quantum mechanics is suggested by the fact that randomness enters both into the initial state and also into the transition from this state to another.

It is possible that some insight may be gained into internal processes in stars and plasmas. In principle, all physical properties of media containing radiating atoms are reflected in the line structure. Thus, a radiating atom can serve as a noninterfering probe conveying significant information regarding pressure, temperature, distribution of molecular speeds, and states of ionization in its surrounding medium. Thus, assuming various probability distributions for parameters of internal processes allows prediction of the spectra to be observed as a result of the action of the medium in terms of the normal or unperturbed spectra and a stochastic kernel or Green's function dependent on the assumed distributions. Successful correspondence of the observed and predicted spectra (line broadening, etc.) would establish the nature of the internal processes.

* This work is a part of the author's Ph.D. dissertation, University of California, Los Angeles, California, 1961.

Other possible applications include cosmic-ray branching processes, biological and economic problems such as birth and death, processes occurring at intervals governed by a probability law (a much desired modification of some earlier treatments), and especially, problems in engineering. Thus, in the latter area we can consider optimum filtering (a generalization of Wiener–Kolmogoroff theory), contributions to guidance, control, or tracking system errors of various unavoidable random effects, the effect of tropospheric fluctuations on the accuracy of location of radio stars, and other problems.

Our primary goal here is the investigation of the properties and use of stochastic operators as a unifying method of dealing with some difficult problems in the hope of gaining further insight into the physical interpretation of the theory in some general physical processes. Later work will attempt calculations, based on a stochastic interpretation, of the Wiener or Feynman integral in strong interaction problems using extensions of this investigation. Also to be emphasized in the following work will be a theory of measurements and problems of wave propagation.

I. STOCHASTIC OPERATIONS ON RANDOM PROCESSES: OPERATOR FORMULATION

1. Development of a Stochastic Green's Function

We are concerned with an investigation of a general class of linear operators we call *stochastic operators*. These are operators involving parameters that are random and require probabilistic treatment. Analogous to the usual representation of operators by matrices, stochastic operators can be represented by stochastic matrices. These are matrices whose elements have a distribution. As randomness vanishes and parameters become deterministic, stochastic operators reduce to ordinary operators.

It is instructive to consider a special subclass of stochastic differential operators given by

$$L = \sum_{\nu=0}^{n} a_\nu(t) \frac{d^\nu}{dt^\nu}$$

and differential equations of the form $Ly = x$, where the nonhomogeneous term $x(t)$, called the input or forcing function, and the operator L, by virtue of the coefficients $a_\nu(t)$, are, in general, stochastic quantities. The converse problem involving the conceptual inversion of the matrix for the random differential operator L is especially interesting. For appropriate operators, we think of y as the result of operating on the input by a stochastic operator symbolically denoted by H or by L^{-1} and otherwise undefined except to say it would include the action of any accompanying boundary conditions. We wish to keep the differential equation formulation in mind in the analysis, although in this section we are primarily interested in general operators H acting on a process x with as little restriction as possible.

In general, stochastic operators arise in the processing of an input quantity (which may itself be a random process) by a linear system, observation, or measurement, in which certain parameters are random or stochastic variables. Problems involving random inputs to fixed linear systems (i.e., where all the a_ν are constants and only x is stochastic) are well understood. A great deal of attention has been devoted also to deterministic time variations in linear systems (linear differential equations with time-varying coefficients). When random variations of system parameters (or of the a_ν in the differential equation example) occur, the analysis is more subtle. The output now for either a deterministic or a random input is a stochastic process. A complete description of the output would involve the determination of all multi-variate probability distributions of the output from similar knowledge of the input. Such a complete description is usually neither possible nor necessary.

Suppose that $x(t)$ is a known input process or the operand. We wish to determine the result of an operation on x by the stochastic operator H. Thus, x might be a communication signal into the proverbial black box which then represents the operator H. The box may be a circuit, filter, servo, observation, experiment, or measurement, or a transmission medium which has random properties. All real media are inhomogeneous, i.e., the mean physical properties vary from one point to another. When, in addition, there is a variance from the mean at each point, the medium is a random inhomogeneous medium. When the properties or parameters of the medium are constants and $x(t)$ is nonstochastic, the output $y = L^{-1}x$ is easily found. When the x is stochastic, and certainly when the properties of the medium vary randomly with time, we can only hope to obtain various statistical measures or estimates of y in terms of similar measures of x. If the inverse H of the stochastic matrix L exists, we have the general problem again given by $y = Hx$ (where we emphasize the product is not multiplicative but denotes the action of the stochastic operator H on the process x) and the differential equation could be solved as well as the general problem Hx. However, the inversion of a stochastic matrix is clearly to be avoided. Since we are asking only for an appropriate statistical measure

of y rather than y itself, we may be able to solve the differential equation at the same time as the general problem without the inversion. A well-known and convenient statistical measure is the power spectral density or power spectrum, or by the Wiener-Khinchin theorem, the autocorrelation function.[1] Thus, if $\Phi_x(f)$ is the power spectrum of the input and $\Phi_y(f)$ is the power spectrum of the output, we would like to obtain Φ_y in terms of Φ_x. Such an expression can be obtained as an integral equation whose kernel K_H involves the random parameters of the operator H. We will then clarify the interpretation of the stochastic kernel K_H and show its derivation from the L as well, providing certain restrictions on L are met. The constant-parameter system is handled easily in terms of Green's functions (weighting functions in control system theory) or frequency response functions and transfer functions, respectively, the Fourier and Laplace transforms of the system Green's function. We now (in the stochastic case) obtain a *stochastic Green's function* that depends on the statistical measure chosen for a solution and which reduces to the ordinary expressions when the parameters are constants or nonstochastic.

Now consider a linear operation $H_{\alpha,\beta,\ldots}$, or H for brevity, on an input process x for which the spectral density Φ_x is known. The operation represents a medium or processing system with parameters $\alpha(t)$, $\beta(t),\cdots$, one or more of which are random. Thus H is determined only by the probability distributions of its parameters, i.e., in the sense of an ensemble of possible operations H_i which have a distribution determined by the parameters. Each of the H_i corresponds to an ordinary nonstochastic matrix. Each member H_i of the ensemble H has associated with it a Green's function $h_i(t,\tau)$ which represents the response at time t of the system H_i to a unit impulse $\delta(t-\tau)$ applied at time τ. More specifically, h_i should be written $h_i(\alpha,\beta,\cdots t;\tau)$ where $\alpha,\beta,\cdots$ are the parameters of H subject to probability laws. The h_i can be obtained either by knowledge of the physical process represented by H_i or by eigenfunction expansion after solving an eigenvalue problem for the deterministic operator H_i. These methods are reviewed later. The response of the H_i system to the input $x(t)$, thinking of x as a defined continuous member of the x process for the moment, can now be given in terms of the Green's function by

$$H_ix(t) = \int_{-\infty}^{\infty} h_i(\alpha,\beta,\cdots t;\tau)x(\tau)d\tau ,$$

or for brevity, where parameter dependence is not used,

$$H_ix(t) = \int_{-\infty}^{\infty} h_i(t;\tau)x(\tau)d\tau , \tag{1}$$

where for causal (physically realizable) systems, $h_i(t,\tau)$ is zero for $t < \tau$. The upper limit can also be written t without change. The implication of the lower limit is that all past values are significant. We can write the spectral density of $H_ix(t)$ by first developing this function in a Fourier integral. Thus assuming stationarity and supposing H_ix to be zero outside some interval $[-T,T]$ then taking the limit as $T \to \infty$

$$H_ix(t) = \int_{-\infty}^{\infty} dfA_i(f)e^{2\pi ift}$$

$$\Phi_{Hi}(f) = \lim_{T\to\infty} \frac{1}{T}|A_i(f)|^2 .$$

Our stationarity assumption can be modified somewhat in that the $A_i(f)$ can depend on T, thus $A_i^T(f)$, as long as the limit exists. The above expression holds for all H_ix which are members of the ensemble Hx. Each H_ix is developed in a Fourier integral as we have shown. The corresponding spectra Φ_{H_i} are averaged over the ensemble. The result which could have been written immediately is $\Phi_{Hx}(f)$, the spectral density of the output. Thus,

$$\Phi_{Hx}(f) = \lim_{T\to\infty} \frac{1}{T} \left\langle \left| \int_{-T}^{T} dtH_i\{x(t)\}e^{-2\pi ift} \right|^2 \right\rangle ,$$

where the average is over the ensembles of H and of x.

$$\Phi_{Hx}(f) = \lim_{T\to\infty} \frac{1}{T} \int_{-T}^{T}\int_{-T}^{T} dt_1dt_2 \times \langle H_i\{x(t_1)\}H_i^*\{x(t_2)\}\rangle e^{-2\pi if(t_1-t_2)} .$$

The quantity in brackets is evidently an autocorrelation function. Let us define the autocorrelation of the Hx as

$$R_{Hx}(\tau) = \langle Hx(t)H^*x(t+\tau)\rangle ,$$

where the random process may be complex and we have limited ourselves to stationary input $x(t)$, though it is possible the operator H may be nonstationary. On using (1), this equation becomes

$$R_{Hx}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\tau_1d\tau_2\langle h(t;\tau_1)h^*(t+\tau;\tau_2)x(\tau_1)x(\tau_2)\rangle .$$

We note the h's correspond to different times t_1 and t_2 (or t and $t+\tau$). Thus we have $h(\alpha_1,\beta_1,\cdots t_1;\tau_1)$ and h^* $(\alpha_2,\beta_2,\cdots t_2;\tau_2)$ so all parameters are different.

[1] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals* (McGraw-Hill Book Company, Inc., 1950).

In many cases of physical interest, the $x(t)$ and the $a_\nu(t)$ are independent. In other words, the input process and the "filtering" process are independent. Transmission of a signal through a randomly varying medium (a stochastic filter) is such a case. The result of the action of the medium on x is the stochastically filtered signal Hx. The term in brackets in the expression for $R_{Hx}(\tau)$ can now be factored. Indicating the autocorrelation for x alone by R_x

$$R_x(\tau_2 - \tau_1) = \langle x(\tau_1)x^*(\tau_2)\rangle ,$$

with averaging obviously over the ensemble of x, we now have

$$R_{Hx}(\tau) = \int d\tau_1 \int d\tau_2 R_x(\tau_2 - \tau_1)\langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle .$$

The corresponding spectral density is

$$\begin{aligned}\Phi_{Hx}(f) &= \int d\tau e^{2\pi i f\tau} R_{Hx}(\tau) \\ &= \int d\tau_1 \int d\tau_2 \int d\tau e^{2\pi i f\tau} R_x(\tau_2 - \tau_1) \\ &\quad \times \langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle \\ &= \int d\tau_1 \int d\tau_2 \int d\tau \int ds e^{2\pi i\tau f} e^{-2\pi i(\tau_2-\tau_1)s} \\ &\quad \times \langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle \Phi_x(s) ,\end{aligned}$$

where $\Phi_x(s)$ is the (power) spectral density of $x(t)$. The last equation is now rewritten in the convenient form

$$\Phi_{Hx}(f) = \int_{-\infty}^{\infty} ds K_H(s,f)\Phi_x(s) , \qquad (2)$$

where the integral kernel or "stochastic kernel"

$$\begin{aligned}K_H(s,f) &= \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} e^{-2\pi i(\tau_2-\tau_1)s} \\ &\quad \times \langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle \qquad (3)\end{aligned}$$

is a spectral representation of the Green's function of the stochastic operator H and may be called a "stochastic Green's function." The form of the kernel or stochastic Green's function depends on the statistical measure chosen for the solution. In solving $Ly = x(t)$, we can express y in terms of x by the use of a Green's function $G(t,\tau)$. However, we are now solving not for y but Φ_y in terms of Φ_x and a "Green's function" appropriate to such measures. This is our kernel K_H or what we call a stochastic Green's function. Since it depends upon the statistical measure, we should properly identify the kernel K_H as a stochastic Green's function for power spectral density measures or a spectral density Green's function.

Now, in principle at least, we can express all the correlations of the output by similar expressions involving correlations of the input for general linear stochastic operations on arbitrary random inputs. But the n-dimensional multivariate characteristic function ϕ_n of the "output" or transformed process can be given in terms of all the correlations of the output, and by the above reasoning can now be given in terms of all the correlations of the input. These, in turn, are clearly determinable from knowledge of all the joint distribution functions of the input process or operand. Finally the n-dimensional joint distribution function for the output is obtained by Fourier inversion of ϕ_n. Thus we have suggested a method, at least in principle, for solving the most general statistical problem.

Equation (2) shows that the power spectrum (or autocorrelation function by use of the Wiener–Khinchin theorem) of the output can be given as a transform of the corresponding measure of the input and a kernel which, as we shall see, involves the probability laws for the parameters. Thus the spectral density of the output is related to the spectral density of the input by a linear operation. The solution of any problem of this type involves finding the appropriate kernel. We write (2) symbolically as

$$\Phi_{Hx} = K_H \cdot \Phi_x$$

understanding K_H to be the appropriate integral operator. We note some properties which would reasonably be expected.

If $x = ax_1 + bx_2 + \cdots$, where the x_i are mutually independent, then $Hx = aHx_1 + bHx_2 + \cdots$ and $\Phi_{Hx} = K_H \cdot \Phi_x = a^2 K_H \cdot \Phi_{x1} + b^2 K_H \cdot \Phi_{x2} + \cdots$. Also if $H = H_1 + H_2$, then $\Phi_{Hx} = (K_{H_1} + K_{H_2}) \cdot \Phi_x$.

Whereas H and the above results are quite general, the L in the differential equation was very special, for convenience, though still an important form. No difficulty in principle would be expected in generalizing to systems of differential equations with many dependent variables (multiple inputs and outputs) or to random fields $V(\mathbf{x},t)$.

To the spectral density derived for the "output" we should add any contributions due to a mean value since the derivation tacitly assumed a zero-mean random process x. When the time average of the process is nonzero, there are singular peaks of the well-known Dirac δ-function type in the spectral density. Thus, to the $\Phi_x(f)$, which represents only the spectrum of what we could call the a c part in electrical language, we should add a term $2\bar{x}^2\delta(f)$. Equivalently, to $\Phi_y(f)$ we add a term $2\bar{x}^2 K_H(0,f)$.

2. Interpretation of the Kernel in Special Cases

Consider first a special case of (2) corresponding to a constant-parameter medium. For such cases (differential equations with constant coefficients and random forcing function), the Green's function depends only on the time interval between application of the impulse and observation of the output. The averaging is obviously unnecessary now, so that (3) can be rewritten

$$K_H(s,f) = \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} e^{-2\pi i(\tau_2-\tau_1)s} \times h(t-\tau_1)h(t+\tau-\tau_2) ,$$

which with appropriate changes of variables becomes

$$K_H(s,f) = \int_{-\infty}^{\infty} d\tau e^{-2\pi i(f-s)\tau} \int_0^{\infty} d\tau_2 e^{2\pi i\tau_2 s} h(\tau_2) \times \int_0^{\infty} d\tau_1 e^{-2\pi i\tau_1 s} h(\tau_1) = |Y(s)|^2 \delta(f-s) ,$$

where

$$Y(s) \doteq \int_0^{\infty} d\tau e^{-2\pi i\tau s} h(\tau) ,$$

i.e., the Fourier transform of the Green's function. Then (2) becomes

$$\Phi_{Hx}(f) = |Y(f)|^2 \Phi_x(f) ,$$

a well-known result.

Returning now to the general stochastic kernel of (2) and proceeding in a formal manner

$$K_H(s,f) = \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} e^{-2\pi i(\tau_2-\tau_1)s} \times \langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle ,$$

at least for stationary kernels

$$K_H = \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} e^{-2\pi i(\tau_2-\tau_1)s} \times R_{h_1,h_2^*}(\tau;\tau_2-\tau_1) = \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 e^{-2\pi is(\tau_2-\tau_1)} \Phi_{h_1h_2^*}(f;\tau_2-\tau_1) ,$$

where $\Phi_{h_1,h_2^*}(f;\tau_2-\tau_1)$ is evidently a cross-spectral density of the h_i defined by

$$\Phi_{h_1h_2^*}(f;\tau_2-\tau_1) = \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} \langle h(\tau;\tau_1)h^*(t+\tau;\tau_2)\rangle = \lim_{T\to\infty} \frac{A_h^T(f;\tau_1)A_h^{*T}(f;\tau_2)}{2T} ,$$

where $A_h^T(f;\tau)$ is the Fourier transform of $h(t;\tau)$

$$h(t;\tau) = \int_{-\infty}^{\infty} df A_h^T(f;\tau) e^{-2\pi ift} .$$

We proceed by computing time averages of representative Green's functions (under the restriction that the processes be ergodic) instead of ensemble averages. The expectation value

$$\langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} d\alpha h(t+\alpha;\tau_1)h^*(t+\tau+\alpha;\tau_2) .$$

Thus, we apply an impulse at τ_1 and make an observation at t for one Green's function and apply an impulse at τ_2 and make an observation at $t+\tau$ for the second Green's function. We then average over a translation of all possible values of the two observation times. The kernel depends upon time in the most general case, while for ergodic operators, the kernel is stationary. For nonstationary problems with time-dependent kernels we need to specify the output for a given interval of time.

Occasionally, the notion of frequency response function is extended to linear time-varying systems by defining:

$$Y(s,t) \doteq \int_0^{\infty} h(t;t-\tau)e^{-2\pi is\tau} d\tau .$$

The general kernel $K_H(s,f)$ becomes

$$K_H = \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} \int_{-\infty}^{\infty} d\tau_2 e^{-2\pi i\tau_2 s} \int_{-\infty}^{\infty} d\tau_1 e^{2\pi i\tau_1 s} \times \langle h(t;\tau_1)h^*(t+\tau;\tau_2)\rangle ,$$

$$= \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} \left\langle \int_{-\infty}^{\infty} d\tau_2 e^{-2\pi i\tau_2 s} h^*(t+\tau;\tau_2) \times \int_{-\infty}^{\infty} d\tau_1 e^{2\pi i\tau_1 s} h(t;\tau_1) \right\rangle ,$$

$$= \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau f} \langle Y^*(s,t+\tau)e^{-2\pi is\tau} Y(s,t)\rangle ,$$

$$= \int_{-\infty}^{\infty} d\tau e^{2\pi i\tau(f-s)} \langle Y^*(s,t+\tau)Y(s,t)\rangle .$$

This is compatible with our definition for the constant-parameter case since the above expression then reduces to

$$\int d\tau e^{2\pi i\tau(f-s)} \langle |Y(s)|^2\rangle = \delta(f-s)|Y(s)|^2 .$$

So far, our stochastic kernel has been quite general. The matrix H has elements with any distribution. However, some statistical properties of

matrices whose elements have normal, or Gaussian, distributions are well known. We can make further identifications by specializing to the case of such distributions (possibly first determining the eigenvalues of the operator then the Green's functions, and consequently the kernel). To see how the distributions enter, it is noted that the Green's functions involve the random parameters $\alpha,\beta,\cdots$, which have distributions $p(\alpha)$, $p(\beta),\cdots$ that must be used in obtaining the average or expectation value involved in the kernel. Thus, for ergodic operators

$$\langle h(\alpha,\beta,\cdots t;\tau_1)h^*(\alpha,\beta,\cdots t+\tau;\tau_2)\rangle = \lim_{T\to\infty}\frac{1}{2T} \times \int_{-T}^{T} du \int_0^\infty d\alpha \int_0^\infty d\beta \int \cdots p(\alpha)p(\beta)\cdots \times h(\alpha,\beta,\cdots t+u;\tau_1)h^*(\alpha,\beta,\cdots t+\tau+u;\tau_2)$$

and we can specialize to certain distributions or even to the case where the distribution is given by a δ function and the parameter is not random.

3. Statistical (Measure) Operations

A class of operations on random processes including, e.g., the ensemble average of a random process, the autocorrelation, mean squared, or higher means, can be called statistical or measure operations. A statistical operator or measure in this sense indicates a (nonstochastic) *deterministic* operation on a *random* process. It is distinguished from ordinary deterministic operations on ordinary functions only in that the operand is now a random process rather than a function, so that the averages are with respect to a random variable. Clearly, any ordinary deterministic operator, e.g., a Laplace (or other) transform should commute with such statistical operators and one could consider a possibility of appropriate (statistical) transforms of random processes since this would just mean the ordinary transform of the result of the statistical operation on the random process—or a transform of a mean.

Stochastic operators present much more difficulty. Let us consider a random function or process $y(t)$. It can be regarded as a generalized vector in an infinite dimensional continuum or a (infinite) continuous (column) matrix, since it consists of an entire ensemble of sample functions or representatives. The elements of the matrix vary according to the distribution to which $y(t)$ is subject. A (stochastic) operator represented by a (stochastic) matrix (with elements subject to an appropriate distribution) now acts on the matrix of the input random process. Clearly, we expect only to find some statistical property of the result. Let us indicate the statistical measure by Γ. Thus, emphasizing the random processes by brackets $\{\ \}$,

$$\{y(t)\} = H\{x(t)\}$$
$$\Gamma\{y(t)\} = \Gamma H\{x(t)\}\ .$$

If we choose spectral density as our statistical measure Γ, we have $\Gamma\{y(t)\} = \Phi_y(f)$ and $\Gamma\{x(t)\} = \Phi_x(f)$. Noting that Γ and H do *not* commute (unless H is deterministic), we have derived Φ_y in terms of Φ_x in the integral relation of Eq. (2), or

$$\Phi_y = \Gamma H\{x(t)\} = \int K_H(s,f)\Phi_x(s)ds\ .$$

When the stochastic operation H and the random process x are statistically independent, the statistical operator Γ must act separately on each, i.e., providing ΓH is defined (and it should be if H is representable by a matrix) and if neither ΓH nor Γx is zero (eliminating a zero-mean random process if Hx does not also have a zero mean) then

$$\Gamma Hx = \Gamma H\cdot\Gamma x$$

This means the appropriate statistical operation over each ensemble, i.e., $\Gamma_H H\cdot\Gamma_x x$. Thus Γx is found by forming $\langle x(t)x^*(t+\tau)\rangle$ then making the transform $\int e^{2\pi i f\tau}\langle\ \ \rangle d\tau$. Similarly ΓH is given by

$$\Gamma H = \int e^{2\pi i f\tau}\langle H(t)H^*(t+\tau)\rangle d\tau\ .$$

Thus we can compute an appropriate autocorrelation and then a transform to get ΓH, or a spectral density of H, if we are given the stochastic matrix of H. Thus we can write

$$\Phi_y = \Phi_H\cdot\Phi_x = \int K_H(s,f)\Phi_x(s)ds\ .$$

If we can find Green's functions, the kernel K_H can be found and Φ_y determined. We can always do this in principle for deterministic operators, as we shall show, and at least for stochastic operators having the ergodicity property, as mentioned earlier. If H is given as a stochastic matrix, K_H and therefore Φ_y correspond to the transform of the autocorrelation of the matrices.

Finally, if H is a *deterministic* operator $\Gamma Hx = H\Gamma x$ or $\Phi_y = H\Phi_x$. Thus if H is the Laplace transform operator, i.e.,

$$H(\ \ ) = \int_0^\infty e^{-st}(\ \ )dt\ ,$$

then

$$\Gamma Hx = \Gamma\int_0^\infty e^{-st}x(t)dt = \int_0^\infty e^{-st}\Gamma x(t)dt = H\Gamma x\ .$$

This clearly is *not* $\Gamma X(s)$ where $X(s)$ is the transform. Γ acts only on the random function $x(t)$. Further even for stationary $x(t)$, Hx or $X(s)$ would not be stationary and $\Gamma X(s)$ would be undefined. But (supposing the statistical measure Γ to be the spectral density) we have

$$H\Gamma x = \int_0^\infty e^{-st}\Gamma x(t)dt = \int_0^\infty e^{-st}\Phi_x(s)dt\,,$$

noting the variable of Φ_x should be the same as the transform variable s. Thus

$$H\Gamma x = \Phi_x(s)\int_0^\infty e^{-st}dt = \frac{\Phi_x(s)}{s}$$

and we see that a statistical measure Γ (in this case the spectral density) of the (Laplace) transform of a random process $x(t)$ is given by a pole of magnitude Γx [the spectral density of $x(t)$] in the transform plane.

4. Determination of the Kernel from Physical Process

In many cases of physical interest (2) can be determined directly from the nature of the process. Thus we have an input $x(t)$ and a "stochastically filtered" or processed output $F_{\alpha,\beta,\ldots}\{x(t)\}$ where the stochastic operator $F_{\alpha,\beta,\ldots}$ depends on the distributions of its parameters $\alpha,\beta,\cdots$. For example, (a) a function $x(t)$ is translated by amounts α_n at intervals of time τ so $F\{x(t)\} = x(t-\alpha_n)$ for $n\tau < t < (n+1)\tau$, (b) $x(t)$ is multiplied by β_n over intervals of length τ. Thus $F\{x(t)\} = \beta_n x(t)$ for $n\tau < t < (n+1)\tau$. These are linear processes and the random variables in question are the α_n and the β_n. The distribution functions of these parameters determine the statistical properties of the process. An example (c), of greater interest arising in an application to be considered is the operation $S\{x(t)\} = x(t_n)$ for $t_n < t < t_{n+1}$ for all n where the intervals $l_n = t_{n+1} - t_n$ are random variables with a distribution $P(l)dl$. Thus S might be called a "random sampling operator" (where the last sample value is held until another is taken). The S operator gives the correct results for the much-used case of regular sampling as randomness vanishes. A closely related example (d) is $P\{x(t)\}$ which samples randomly as before but inverts alternate samples. Thus $P\{x(t)\} = (-1)^n x(t_n)$ for $t_n < t < t_{n+1}$ where the intervals $l_n = t_{n+1} - t_n$ are specified by $P(l)dl$.

The Green's functions for these latter processes (c and d) can be written by inspection

$$s_i(t,\tau) = \textstyle\sum_k \Delta(t_k,l_k;t)\delta(\tau - t_k)$$
$$p_i(t,\tau) = \textstyle\sum_k (-1)^k\Delta(t_k,l_k;t)\delta(\tau - t_k)\,,$$

where

$$\Delta(t_k,l_k;t) = 1 \quad \text{if} \quad t_k < t < t_k + l_k$$
$$= 0 \quad \text{otherwise}\,.$$

The spectral densities for the results of operations on x by S or P are considered in the referenced dissertation in connection with a generalized random-walk problem, first studied jointly with DuBois. We state the results here as examples.

$$\Phi_{Sx}(f) = \mathrm{Re}\left\{\frac{1-z(f)}{2\pi^2 f^2 \bar{l}}\int_{-\infty}^{\infty} ds\Phi_x(s-f) \times [1-z(s-f)]\sum_0^\infty z^n(s)\right\}$$
$$\Phi_{Px}(f) = \mathrm{Re}\left\{\frac{1-z(f)}{2\pi^2 f^2 \bar{l}}\int_{-\infty}^{\infty} ds\Phi_x(s-f) \times [1+z(s-f)]\sum_0^\infty (-1)^n z^n(s)\right\},$$

where $z(f)$ is the characteristic function of the distribution for the sampling intervals $P(l)$, i.e.,

$$z(f) = \int_0^\infty dlP(l)e^{-2\pi i f l}$$
$$\bar{l} = \int_0^\infty lP(l)dl = -\frac{1}{2\pi i}z'(0)\,,$$

relations which are clearly of the same form as (3) with rather complicated kernels dependent on the probability laws [in this case just $P(l)$] for the parameters which are random. Evaluation of the kernel in terms of distributions of the random parameters is obtained directly from the nature of the process on the input, i.e., from Hx.

5. Stochastic Green's Functions for Correlation Measures of Input and Output

The stochastic Green's functions or stochastic kernels for other statistical measures of input and output can be found in the same way as carried out for spectral density measures. Some of these kernels, e.g., those for correlation functions, are particularly valuable, both because we frequently want to determine the correlation matrix for a random process and because of our suggestion for determination of the nth joint distribution function, if we can first find all the output correlations in terms of the input correlations.

The kernel for the autocorrelation measure of input and output for the stationary case can be seen from the derivation of K_H in Sec. I. Thus, the relation of the autocorrelations R_{Hx} of the output and R_x of the input for an operation H is

$$R_{Hx}(\beta) = \int_{-\infty}^{\infty} d\sigma G_H(\beta,\sigma)R_x(\sigma)\,,$$

where the kernel G_H is given by

$$G_H(\beta,\sigma) = \int_{-\infty}^{\infty} d\tau \langle h(t,\tau)h^*(t+\beta,\tau+\sigma)\rangle .$$

Transformation of R_{Hx} and R_x to spectral densities again gives the relation involving the spectral density kernel $K_H(s,f)$. Because G_H is simpler than K_H, it is usually desirable to work with correlations although this investigation has emphasized spectra.

6. Nonstationary Processes[2]

If we had not restricted ourselves to stationary processes in writing $R_{Hx}(\tau)$ (early in Sec. I), we would instead write $R_{Hx}(t_1,t_2)$ or simply $R(t_1,t_2)$ given by

$$R(t_1,t_2) = \iint \langle h_i(t_1,\sigma)x(\sigma)h_i(t_2,\tau)x(\tau)\rangle d\sigma d\tau$$
$$= \iint R_x(\sigma,\tau)\langle h_i(t_1,\sigma)h_i(t_2,\tau)\rangle d\sigma d\tau ,$$

where $R_x(\sigma,\tau)$ is the correlation of the input. Thus,

$$R(t_1,t_2) = \iint R_x(\sigma,\tau)H(t_1,t_2,\sigma,\tau)d\sigma d\tau ,$$

where $H(t_1,t_2,\sigma,\tau)$ can be considered a kernel for the autocorrelation where the processes need not be stationary. When the processes are stationary,

$$R_x(\sigma,\tau) = R_x(\sigma-\tau) ,$$
$$h_i(t_1,\sigma) = h_i(t_1-\sigma) ,$$
$$h_i(t_2,\tau) = h_i(t_2-\tau)$$

and H becomes $H(t_1-\sigma,t_2-\tau)$ a function of 2 variates as in $K_H(s,f)$. We are still assuming independence of the system and the input. The output is stationary and we can write our spectral density expression by appropriate transformation.

7. Mapping of Spaces

If the measure chosen for $x(t)$ and $y(t)$ is the spectrum Φ, we have

$$\Phi_y(f) = \int K(s,f)\Phi_x(s)ds .$$

If the measure chosen is the autocorrelation we have

$$R(\beta) = \int G(\beta,\sigma)R_x(\sigma)d\sigma .$$

A mathematically more elegant formulation is to define a general statistical measure μ.[3] Thus the measure of y is $\mu(y)$ and the measure of x is $\mu(x)$. Now $\mu(y)$ may be $\Phi(f)$ or $R(\tau)$, for example. It means a chosen measure defined over the space of the random function y. Similarly, $\mu(x)$ is defined over the space of $x(t)$ and may, for example, be the spectrum $\Phi_x(s)$. Now we can define

$$\Phi_x(s)ds = d\mu(x)$$

and write (1) as

$$\mu(y) = \int K(y,x)d\mu(x) .$$

This form is now appropriate to any chosen measure operation. The kernel $K(y,x)$ maps from the space of x to the space of y and depends on the chosen μ.

A stochastic process $X(t)$,[4] can be described in terms of n-dimensional random variables $X(t_1)$, $X(t_2),\cdots X(t_n)$ by means of their families of distribution functions $F_n(t_1;X_1,t_2;X_2,\cdots,t_n;X_n)$. However, such a law will be meaningless on the quantum level since it implies the observation of a physical phenomenon at an instant determined by a point on a time scale, or generalizing immediately to random fields for precise measurements of position, momentum, or other variables. Such precise or instantaneous observations are inconsistent with basic postulates so a reinterpretation becomes necessary. However, it is not essential to consider random processes as families of random variables depending upon t (or $\mathbf{t}$). Instead we think of X as an element in a function space. Each realization of the process $X(t)$ becomes one point in the function space. We define a general random element X with values in a space χ in the following way. Let $x(u)$ be a mapping from a measurable space of elementary events U into the space χ. The random element X is the result of the mapping. The elements of χ are the values which X can take. Fortet[5] gives some good examples. If X is a random variable, χ is the space R of real numbers. If $\underset{\sim}{X}$ is a k-dimensional random variable, then χ is a k-dimensional Euclidean space. If X is a Gaussian random function of t on an interval (t_1,t_2) (with continuous covariance) then χ may be either the Hilbert space of those functions on (t_1,t_2) whose square is integrable or the space of all functions on (t_1,t_2).

Now we can extend the definition of stochastic operators to abstract spaces. Thus, the stochastic operator L acting on the random function X can be defined as the transformation which maps the space χ into a space $\mathfrak{X}$ representing values which LX can take. Any operator satisfies the definition so far;

[2] This treatment was suggested by Professor Balakrishnan, of the University of California at Los Angeles.

[3] Called Γ in the section on statistical measure operations.

[4] We can generalize to a random field by writing $X(t,\mathbf{r})$ or simply letting t be a vector, i.e., X can be a function of several variables.

[5] R. Fortet, *Recent Advances in Probability Theory* (John Wiley & Sons, Inc., New York, 1958).

now we include linear transformations which themselves require statistical treatment in terms of realizations such as we gave $X(t)$. A single realization of a stochastic operator yields an ordinary operator or nonrandom matrix. It is possible that a formulation of quantum theory using stochastic operators may couple the advantages of the Feynman formulation with calculable expressions and a more natural appearance of probabilities. This work is intended to lay the foundation for later use in this connection. The extension will be of the nature of a general transformation theory involving topological spaces and functional analysis.

8. Some General Mathematical Properties of Kernels and Operators[6]

We note in passing some general mathematical properties of the kernel and its relations to stochastic operations and statistical measure operations. If $y = Hx$, where H is a stochastic operation on the stochastic process or random function x and we take the statistical measure μ of both sides of this equation, we have

$$\mu(y) = \mu(Hx) = \int K(y,x)d\mu(x) \doteq K\mu(x)$$

(it being assumed that μ does not assign finite measure to a point). In order that the integral operator K exists, μ must have certain properties with respect to H. From the above equation we can write the *operator* equation

$$\mu H - K\mu = 0 .$$

Let us assume that the quantity $H\mu$ is operationally defined in some sense. We have already seen in the section on measure operations that μ does not commute with stochastic operations, i.e.,

$$\mu H \neq H\mu$$

Since we have assumed that $H\mu$ is defined, we may define the operator ζ by

$$\mu H = H\mu + \zeta\mu .$$

Substituting this into the equation $\mu H - K\mu = 0$, we have

$$(H + \zeta - K)\mu = 0 .$$

Thus, under the assumption that $H\mu$ exists, we have

$$K = H + \zeta + \sigma ,$$

where σ is any operator such that

$$\sigma\mu = 0 .$$

Conversely, if K exists and $H\mu$ is defined, then the equation $\mu H - K\mu = 0$ is satisfied if and only if there exists an operator ζ such that

$$\mu H = H\mu + \zeta\mu .$$

Combining the above considerations, we obtain the following result:

Theorem. If $H\mu$ is defined, there exists an operator K which satisfies the operator equation

$$\mu H - K\mu = 0$$

if and only if there exists an operator ζ such that

$$\mu H = H\mu + \zeta\mu .$$

If these conditions are satisfied, then K is given by

$$K = H + \zeta + \sigma ,$$

where σ is any operator such that $\sigma\mu = 0$.

Now assume that σ is zero so $H + \zeta$ represents K. The procedure is clear in principle. For a given stochastic operation and chosen measure operation, a kernel K exists if $\mu H - K\mu$ is a set of measure zero over the space of x on which H acts. The commutator of μ with H determines ζ. Then $K = H + \zeta$.

If we consider μ and H as operators each with a unique set of eigenvectors and suppose these sets are nonparallel, i.e., μ and H do not commute, then ζ is orthogonal to H, i.e., the eigenvectors of ζ are orthogonal to the eigenvectors of H. We are defining an operator ζ such that the eigenvectors of $H + \zeta$ commute with K. For H and μ such that ζ is sufficiently simple, it is conceivable that one could determine K from $H + \zeta$.

Now consider the inverse problem $Ly = x$ where we are particularly, but not exclusively, interested in the (unbounded) differential operator. Again we suppose that $\mu(x)$ is a well-defined measure given for x and we ask for $\mu(y)$. Write $\mu(Ly) = \mu(x)$. L and y are not statistically independent, so μ does not act separately on L and y as in the case Hx. Again the measure operation μ does not commute with the stochastic operation L. Suppose the commutation relation is

$$\mu L = L\mu + \chi\mu ,$$

where again χ may be, and in general is, quite complicated. Now supposing $\mu(y)$ exists,

$$L\mu(y) + \chi\mu(y) = \mu(x)$$

or

$$LK\mu(x) + \chi K\mu(x) = \mu(x) ,$$

[6] Not used in further development. This section is of the nature of a mathematical appendix concerning meaning of operations and is based on a discussion with D. Edelen at Rand.

and we can write the operator relation

$$(L + \chi)K = 1 .$$

Thus, given the measure operator μ and the stochastic operator L, the operator K is defined and exists when χ exists or when $\mu(y)$ and $\mu(x)$ exist. However this is not a method of calculating K since we do not know how to calculate the inverse of a stochastic matrix. We may also note that even if $\mu(x)$ exists, $\mu(y)$ may not exist. L is an unbounded operator (for the differential operators) which may be nonmeasure-preserving. The space of y must be suitably restricted. If a measure on x space is mapped into the same measure on y space, then it is clear that not all measures will work for an arbitrary L.

This shows $(L + \chi)\ (H + \zeta) = 1$ or $H + \zeta$ is inverse to $L + \chi$, but we started with the supposition that H was inverse to L by writing $y = Hx$ as the solution for $Ly = x$. This implies that $\chi H + L\zeta + \chi\zeta = 0$ or that χ and ζ are related for a particular L and its inverse H.

We will now generalize our remarks in the following manner. We have seen that we can write

$$\Phi_y(f) = \int ds K_H(s,f)\Phi_x(s)$$

$$\doteq W(s,f)\Phi_x(s)$$

defining $W(s,f)$ to mean the integral operator $\int ds K_H(s,f)$. The above expression is equivalent to

$$\Gamma Hx = W(s,f)\,\Gamma x ,$$

where Γ represents the statistical measure operation of forming the spectral density. To keep track of the arguments f or s we use subscripts on the Γ operator. Thus, $\Phi_y(f) = \Gamma_f Hx$ and $\Phi_x(s)[= \Gamma_s x$. Now

$$\Gamma_f Hx = W(s,f)\,\Gamma_s x = \Gamma H \cdot \Gamma_f x$$
$$= \Gamma H\delta(s - f)\,\Gamma_s x$$

so that

$$W(s,f) = \Gamma H\delta(s - f) .$$

Also, knowing ΓH is a noncommuting pair, we write

$$\Gamma_f H = H\Gamma_f + \eta\Gamma_f$$
$$\eta\Gamma_f = \Gamma_f H - H\Gamma_f$$
$$= \Gamma H\delta(s - f)\,\Gamma_s - H\delta(s - f)\,\Gamma_s$$
$$= [W(s,f) - H\delta(s - f)]\,\Gamma_s .$$

Thus, $\eta\Gamma_f$ is an essentially singular operation. η is related to our kernel operation $W(s,f)$ by an additive Dirac matrix. We have shown that η exists and shown its specific structure. If Γ were to represent an ordinary averaging operation and H a stochastic operation, then ΓHx is a defined operation. $H\Gamma x$ is undefined, but the Dirac function behavior of η is just such as to eliminate the point in space from which the indeterminism results.

We can state now the following as a theorem:

Theorem. There exists a kernel $K(s,f)$ or the integral operator $W(s,f) = \int_{-\infty}^{\infty} ds K_H(s,f)$ if and only if there exists an η satisfying the commutation relation $\Gamma H - H\Gamma = \eta\Gamma$, such that

$$W(s,f) = \Gamma H\delta(s - f) + \chi ,$$

where χ is an arbitrary operator such that $\chi\Gamma x = 0$, i.e., χ annihilates Γ.

We have shown $K(s,f)$ exists under the assumption $\Gamma Hx = \Gamma H \cdot \Gamma x$. Now one could consider all Γ for example that result in $\Gamma Hx = \Gamma H \cdot \Gamma x$ plus some other term and again find a necessary and sufficient condition for existence of a kernel.

9. Physical Measurements and Estimates of Statistics

With *a priori* knowledge of probability distributions of given random processes we have calculated various statistical measures (moments, characteristic functions, spectral densities, etc.,) for these distributions.

In physical measurements of random processes with unknown distributions (where there is no possibility of taking advantage of the central limit theorem) a question of great importance we may ask, is whether from a set of measurements we can now determine some of the statistical properties (mean, variance, etc. . . .).

Suppose we have available N measurements of a basic experiment defining a process or N observations or measurements of the value of a sample function of a given random process at N different instants of time. This set of N values is a sample point in a N-dimensional space characterizing the entire experiment. Clearly, this single point will not uniquely determine the statistical property we seek. In this situation the statistician speaks of making an *estimate* of the particular statistical property by finding a function or *statistic* of the set of results or values which gives a reasonably close estimate of the statistical property for the process. This is called sampling theory and is discussed by Davenport,[7] Cramer,[8] and others. It is clear from the theory and our own results on the random sampling example,

[7] W. B. Davenport, and W. L. Root, *An Introduction to the Theory of Random Signals and Noise* (Massachusetts Institute of Technology, Cambridge, Massachusetts, 1958).

[8] H. Cramer, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946), Part III.

that we can now give more generalized sampling theorems for situations in which it may matter that the N instants of time at which measurements are made may not be regularly spaced but are subject to a distribution.[9] Such a case might occur, for example, if the observations were to be made whenever an incoming particle triggered the observing apparatus.

The application of the theory to the statistical description of mechanical systems of macroscopic dimensions is a clear possibility. We can assume a stochastic process corresponding to the measurement of an n-tuple of macroscopic observables. Then a theory of repeated observations or measurements can be made using stochastic operators, where some parameter of the observation is unknown and can only be given a distribution.

II. THE INVERSE PROBLEM: STOCHASTIC DIFFERENTIAL EQUATIONS

1. Differential Operators

We have discussed at some length in Part I various statistical measures for the action of a linear stochastic operator on a random process, or equivalently, a linear (stochastic) transformation of a random process. A particular measure of interest, the spectral density measure, involves a stochastic kernel or spectral density Green's function K_H for a stochastic operator H. This formulation is useful in some problems involving the statistical optimization of complex systems and in the synthesis of models for some physical phenomena. In other physical problems, however, the convenient formulation involves a "stochastic differential equation" rather than a given operator. Generally, linear operators, linear transformations, linear filters, linear networks, and linear control systems are equivalent mathematical systems describable by systems of linear integro-differential equations. The forcing functions are the inputs. The outputs or dependent variables appear as the operands. For simplicity, we consider a single input, single output equation in the form

$$Ly = x$$

where

$$L = \sum_{\nu=0}^{n} a_\nu(t) \frac{d^\nu}{dt^\nu}$$

and both $a_\nu(t)$ and $x(t)$ are stochastic. Samuels,[10] Sundstrom,[11] and others have pointed out that such problems arise in the study of linear servomechanisms in which several parameters are undergoing noise modulations[12,13] and in certain propagation problems in stochastically varying media[14–17] or media with random inhomogeneities.

We now consider the equation $Ly = x$, with L the previously defined stochastic operator and $x(t)$ a random process. Defining H as the operator inverse to L, we investigate the possibility of determining the kernel K_H for H by examining L and avoiding the problem of inversion, i.e., not determining H. We assume reasonable restrictions on L.

1. The inverse must exist. Thus L must be nonsingular, i.e., the determinant of the matrix corresponding to L must never become zero as the elements vary according to some distribution.

2. K_H is stationary or, at least, the coefficients in L can be translated, i.e., have stationary properties.

Now, differential equations can be solved in our sense, i.e., in terms of a statistical measure, where the coefficients are stochastic as well as where the coefficients can be considered to be chosen from random processes. Our stationarity condition means simply that the probability distributions of the random parameters are such that the stochastic coefficients (time functions) in the differential equation describe stationary processes during a time interval of interest, i.e., the statistical characteristics during the interval are constants. In order to have a full description of the processes, the time interval must be at least so long that the correlation between values at the limit points can be neglected. Actually, wide-sense stationarity is sufficient for the coefficients.

2. Evaluation of Stochastic Green's Function for a Differential Operator

A series of papers by Sundstrom[18] leads to the following approach for expressing the kernel in terms of the coefficients of the differential equation. Treating x not as a random process but as a defined continuous member of the x process in the interval $[-T/2, T/2]$, and similarly treating y and a_ν, we

[9] Since the term random sampling is already used by statisticians to indicate that the sample is chosen from a general population in a random manner, it might be better to call this stochastic sampling.

[10] J. C. Samuels, *Transactions of the 1959 International Symposium on Circuit and Information Theory, Los Angeles, June 16–18, 1959*. IRE Trans. Circuit Theory, **CT-6** (Special Supplement) (1959).

[11] M. Sundstrom, Arkiv Mat. **2**, 52 (1951).

[12] A. Rosenbloom, Ph. D. Thesis, UCLA, 1954 (unpublished).

[13] J. F. Buchan, and R. S. Raven, IRE (Wescon) Convention Record, Part 4, August 1957 (unpublished).

[14] P. G. Bergman, Phys. Rev. **70**, 486 (1946).

[15] D. Mintzer, J. Acoust. Soc. Am. **26**, 186 (1954).

[16] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill Book Company, Inc., New York, 1960).

[17] F. Villars, and V. F. Weisskopf, Phys. Rev. **94**, 232 (1954).

[18] M. Sundstrom, Tech. Notes TN 45, 46, 47, 48, 49, 50, 51, Royal Institute of Technology, Stockholm, Sweden (unpublished).

define Fourier transforms of the "input" x, the "output" y, and the coefficients a_ν.

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{2\pi ift}df$$

$$y(t) = \int_{-\infty}^{\infty} Y(f)e^{2\pi ift}df$$

$$a_\nu(t) = \int_{-\infty}^{\infty} A_\nu(f)e^{2\pi ift}df\,. \tag{4}$$

Now if $y(t)$ is differentiated ν times, we have

$$y^{(\nu)}(t) = \int_{-\infty}^{\infty} (2\pi if)^\nu Y(f)e^{2\pi ift}df\,.$$

Substituting in the differential equation

$$\sum_{\nu=0}^{n} a_\nu(t)y^{(\nu)}(t) = x(t)\,.$$

Multiplying by $e^{-2\pi ist}$ and integrating from $-T/2$ to $T/2$, we obtain:

$$\int_{-\infty}^{\infty} df\mathfrak{A}(f,s)Y(f) = X(s) \tag{5}$$

as a formal relation between input and output transforms where:

$$\mathfrak{A}(f,s) = \sum_\nu (2\pi if)^\nu A_\nu(s-f) \tag{6}$$

However, it is clear that we need an expression for $Y(f)$ in terms of $X(s)$, since ensemble averages to be taken later must be separable and, while the x can be statistically independent of the a_ν, the output y cannot, except in a trivial case, and except for one special case of some interest considered by Samuels.[19] He analyzes random linear systems containing one or more nonindependent parameters under the restriction that the parameter processes and the solution or output process have very widely separated spectra. He considers an equation of our form $Ly = x$ with the $a_\nu(t)$ and the $x(t)$ as prescribed random functions, supposing as in our treatment that $x(t)$ is statistically independent of any $a_\nu(t)$ while the $a_\nu(t)$ may or may not be correlated with each other. Following his treatment we assume the mathematical expectation[20] $\langle a_\nu(t)\rangle$ of each coefficient a_ν exists and write

$$a_\nu(t) = \langle a_\nu(t)\rangle + \alpha_\nu(t)$$

so that α_ν represents the randomly fluctuating part of $a_\nu(t)$. The differential equation can now be written

$$Hy = \sum_0^n \langle a_\nu(t)\rangle \frac{d^\nu y}{dt^\nu} = x(t) - \sum_0^n \alpha_\nu \frac{d^\nu y}{dt^\nu}\,.$$

Now the left side can again be viewed as an operator acting on y, but this is now a nonrandom operator H, and a Green's function can immediately be written, say $h(t,\tau)$, for the left side alone. Now the solution is

$$y(t) = \int_{t_0}^{t} h(t,\tau)x(\tau)d\tau - \int_{t_0}^{t} h(t,\tau)\sum_{\nu=0}^{n}\frac{d^\nu y(\tau)}{d\tau^\nu} \times \alpha_\nu(\tau)d\tau + \sum_{\nu=1}^{n} c_\nu\phi_\nu(t)\,,$$

where the ϕ_ν are a fundamental set of independent solutions of the homogeneous nonrandom equation $Hy = 0$ and the c_ν are arbitrary constants. The upper limit can be ∞ for a realizable system. The lower limit is a matter of memory. Using t_0 instead of $-\infty$ implies simply that all earlier values are not significant. Considering infinite limits

$$y(t) = F(t) - \int_{-\infty}^{\infty} K(t,\tau)y(\tau)d\tau\,,$$

where

$$K(t,\tau) = h(t,\tau)\sum_{\nu=0}^{n}\alpha_\nu(\tau)\frac{d^\nu}{d\tau^\nu} = \sum_{\nu=0}^{n}(-1)^\nu\frac{\partial^\nu}{\partial\tau^\nu}[\alpha_\nu(\tau)h(t,\tau)]\,,$$

$$F(t) = \int_{-\infty}^{\infty} h(t,\tau)x(\tau)d\tau + \sum_{\nu=1}^{n} c_\nu\phi_\nu(t)\,.$$

Now multiply $y(t_1)$ and $y^*(t_2)$ and average

$$\langle y(t_1)y^*(t_2)\rangle = \langle F(t_1)F^*(t_2)\rangle - \int_{-\infty}^{\infty}\langle K(t,\tau)F^*(t_2)y(\tau)\rangle d\tau - \int_{-\infty}^{\infty}\langle K^*(t_2,\tau)F(t_1)y^*(\tau)\rangle d\tau + \iint_{-\infty}^{\infty}\langle K(t_1,\tau_1)K^*(t_2,\tau_2)y(\tau_1)y^*(\tau_2)\rangle d\tau_1 d\tau_2\,.$$

Assume the α_ν are stationary and ergodic and the $\langle a_\nu\rangle$ are constants so $h(t,\tau)$ is $h(t-\tau)$. Then it is true that $y(t)$ and $\alpha_\nu(t)$ $y(t)$, etc., are stationary and ergodic. [Also, Samuels states if the $\langle a_\nu(t)\rangle$ approach constants as $t\to\infty$ then the products in question are "almost ergodic" for large t.] We then assume both K and y are ergodic. Now we replace ensemble averages by time averages,[21] e.g.,

[19] J. C. Samuels, *Transactions of 1959 International Symposium on Circuit and Information Theory, Los Angeles, June 16–18, 1959*. IRE Trans. Circuit Theory, **CT-6** (Special Supplement) (1959).

[20] Mathematical expectation or statistical average or mean or ensemble average of a continuous random variable $g(x)$ is

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x)\,p(x)\,dx = \langle g(x)\rangle\,.$$

[21] Time averages are indicated by a bar over a single letter without exponents. For more complicated expressions we will use the symbol $[\ ]_{av}$.

$$E\{G_1(\xi)G_2(\xi)\} = \lim_{x\to\infty}\frac{1}{2x}\int_{-x}^{x} G_1(\xi)G_2(\xi)d\xi = \overline{G}_1\overline{G}_2$$

if the rates at which the functions change are reasonably far apart, e.g., if G_1 is a rapidly changing quantity and G_2 is slowly varying, in which case the plot of the autocorrelation of G_2 is wide and that of G_1 is very narrow, so the spectral densities are well separated narrow peaks. (G_2 close to zero frequency and G_1 farther out.) This is also shown by integrating by parts assuming one rate, say $G_2'(\xi) \simeq 0$. Now we can make the separations

$$\langle K(t,\tau)F^*(t_2)y(\tau)\rangle = \langle K(t,\tau)\rangle\langle F^*(t_2)y(\tau)\rangle$$
$$\langle K^*(t_2,\tau)F(t_1)y^*(\tau)\rangle = \langle K^*(t_2,\tau)\rangle\langle F(t_1)y^*(\tau)\rangle$$
$$\langle K(t_1,\tau_1)K^*(t_2,\tau_2)y(\tau_1)y^*(\tau_2)\rangle = \langle K(t_1,\tau_1)K^*(t_2,\tau_2)\rangle\langle y(\tau_1)y^*(\tau_2)\rangle .$$

Substituting and noting $\langle K\rangle = 0$, since $\langle\alpha_\nu\rangle = 0$,

$$\langle y(t_1)y^*(t_2)\rangle = R_y(t_1,t_2)$$
$$\langle F(t_1)F^*(t_2)\rangle = R_F(t_1,t_2) ,$$

we now have

$$R_y(t_1,t_2) = R_F(t_1,t_2) + \iint_{-\infty}^{\infty}\langle K(t_1,\tau_1)K^*(t_2,\tau_2)\rangle R_y(\tau_1,\tau_2)d\tau_1 d\tau_2 .$$

The above kernel is given by

$$\langle K(t_1,\tau_1)K^*(t_2,\tau_2)\rangle = \sum_{j,k=0}^{n}(-1)^{j+k}\frac{\partial^{j+k}}{\partial\tau_1^k\partial\tau_2^j} \times \{\rho_{kj}(\tau_1,\tau_2)H(t_1,\tau_1)H(t_2,\tau_2)\} ,$$

which is a convenient form when the various cross correlations ρ_{kj} are given. If the cross-spectral densities are specified and $\langle a_\nu(t)\rangle$ are constants, a more convenient form is obtained by taking a double Fourier transform of $R_y(t_1,t_2)$ thus defining

$$\iint_{-\infty}^{\infty} e^{i(\zeta_1 t_1+\zeta_2 t_2)}R_y(t_1,t_2)dt_1dt_2 = \Phi_y(\zeta_1,\zeta_2)$$

with inverse

$$R_y(t_1,t_2) = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} e^{-i(\zeta_1 t_2+\zeta_2 t_2)}\Phi_y(\zeta_1,\zeta_2)d\zeta_1 d\zeta_2 .$$

We can now write

$$\Phi_y(\zeta_1,\zeta_2) = \Phi_F(\zeta_1,\zeta_2) + \sum_{k,j}^{n}\frac{(-1)^{k+j}}{4\pi^2}\iint_{-\infty}^{\infty}\Phi_H(\zeta_1) \times \Phi_H(\zeta_2)\beta_1^k\beta_2^j\Phi_{\rho_{kj}}(\zeta_1-\beta_1;\zeta_2-\beta_2)\Phi_y(\beta_1,\beta_2)d\beta_1 d\beta_2 ,$$

where $\Phi_{\rho_{k,j}}$ is the cross-spectral density of the coefficients. If the α_ν are wide-sense stationary,

$$\Phi_{\rho_{k,j}}(\zeta_1-\beta_1;\zeta_2-\beta_2) = 2\pi\Phi_{\rho_{kj}}(\zeta_2-\beta_2)\delta(\zeta_1-\beta_1+\zeta_2-\beta_2) .$$

Now we have a second-order linear integral equation which can be solved in closed form. Samuels develops a mean square stability theory and considers some special cases, e.g., an *RLC* circuit with randomly varying capacitance. He applies Kirchoffs laws to get a differential equation then considers mean-square stability, i.e., regions where

$$\lim_{t\to\infty}\langle y^2(t)\rangle < M ,$$

where M is a finite constant. Mean-square capacity deviation is plotted against damping of the circuit to show regions of stable operation.

These results, though interesting, are not sufficient for our purpose since they apply to a special case where the ensemble averages can be separated on the basis of rate. Furthermore, we would really like a solution for the output in terms of the input (see Sec. I) and the parameter variation rather than in terms of an integral involving the output again. Therefore, we return to Eq. (5).

We define a new function $B(\sigma,s)$ by the following orthogonality requirement

$$\int_{-\infty}^{\infty} ds\mathfrak{A}(f,s)B(\sigma,s) = \delta(f-\sigma) . \tag{7}$$

From (5) and (7) we immediately obtain

$$Y(f) = \int_{-\infty}^{\infty} dsB(f,s)X(s) . \tag{8}$$

The power spectrum (of the representative member of the ensemble we have chosen) is then given by

$$G_y(f) = \lim_{T\to\infty}\frac{\mathscr{A}}{T}\iint_{-\infty}^{\infty} d\sigma dsB(\sigma,f)B^*(s,f)X(\sigma)X^*(s)$$

The spectrum of the *process is* $\langle G_y(f)\rangle$ or

$$\Phi_y(f) = \lim_{T\to\infty}\frac{2}{T}\iint_{-\infty}^{\infty}\langle B(\sigma,f)B^*(s,f)X(\sigma)X^*(s)\rangle d\sigma ds.$$

But B depends only on the coefficients a_ν not on the input x so that the parameter processes and the input process are statistically independent. Thus,

$$\Phi_y(f) = \lim_{T\to\infty}\frac{2}{T}\iint_{-\infty}^{\infty}\langle B(\sigma,f)B^*(s,f)\rangle \times\langle X(\sigma)X^*(s)\rangle d\sigma ds ,$$

$$\Phi_y(f) = \lim_{T\to\infty}\frac{2}{T}\iint_{-\infty}^{\infty}\langle B(\sigma,f)B^*(s,f)\rangle \times\left\langle\int_{-T/2}^{T/2}x(t)e^{-2\pi i\sigma t}dt\int_{-T/2}^{T/2}x^*(t)e^{2\pi i s t}dt\right\rangle d\sigma ds,$$

$$\Phi_y(f) = \lim_{T\to\infty}\frac{2}{T}\iint_{-\infty}^{\infty}\iint_{-T/2}^{T/2} d\sigma dsdudv\langle B(\sigma,f)B^*(s,f)\rangle \times\langle x(u)x^*(v)\rangle e^{2\pi i(sv-\sigma u)} .$$

If the input is stationary ($v = u + \tau$) this becomes

$$\lim_{T\to\infty} \frac{2}{T} \int \cdots \int d\sigma ds du d\tau \langle B(\sigma,f)B^*(s,f)\rangle \times \langle x(u)x^*(u+\tau)\rangle e^{2\pi i(s-\sigma)u} e^{2\pi i s\tau} .$$

But, $\langle x(u)x^*(u+\tau)\rangle = R_x(\tau)$ and $\int R_x(\tau)e^{2\pi i s\tau}d\tau = \Phi_x$ in the limit $T \to \infty$. Thus we can write our earlier result from Sec. I.

$$\Phi_y(f) = \int ds K_H(s,f)\Phi_x(s)$$

with K_H now defined by

$$K_H(s,f) = \int_{-\infty}^{\infty} d\sigma \int_{-\infty}^{\infty} du \langle B(\sigma,f)B^*(s,f)\rangle e^{2\pi i(s-\sigma)u}$$
$$= \int d\sigma \langle B(\sigma,f)B^*(s,f)\rangle \int du e^{2\pi i(s-\sigma)u}$$
$$= \int d\sigma \delta(s-\sigma)\langle B(\sigma,f)B^*(s,f)\rangle$$

or

$$K_H(s,f) = \langle |B(s,f)|^2\rangle , \tag{9}$$

so that the stochastic kernel has been expressed in terms of the coefficients of the differential equation. While this result establishes the existence of the connection we were seeking in a surprisingly simple relationship, it can only be useful in solving problems or specific equations if we can evaluate it, using the definitions (6) and (7). Now the difficulties become clear, for if $\mathfrak{a}$ and B are represented by matrices, these matrices must be orthogonal and the two are obviously not independent so we cannot make separations in any ensemble averaging. Thus, we have not shown a prescription for the solution of differential equations unless a method can be shown for finding the ensemble mean given by (9) in terms of the given statistics of the $a_\nu(t)$. Let us restate our problem in a convenient way. We wish to solve for B in the integral equation

$$\int_{-\infty}^{\infty} ds \mathfrak{a}(f,s)B(\sigma,s) = \delta(f-\sigma) ,$$

considering $\mathfrak{a}(f,s)$ to be a known kernel, is, an equation of the form

$$I\theta = x ,$$

where x is the Dirac δ function, θ is the unknown function $B(\sigma,s)$ and I is the integral operator with a symmetric kernel $\mathfrak{a}(f,s)$. In the study of transport phenomena (Boltzmann equation) we have just such an inhomogeneous integral equation[22] with a known symmetric kernel which is positive definite, a known inhomogeneous term x, and an unknown function θ. Kohler[23] has shown that the Boltzmann equation is equivalent to a variational principle for which the variational function is simply the inner or scalar product $(\theta,I\theta)$ subject to the subsidiary condition $(\theta,I\theta) = (\theta,x)$. We now show that this statement applies for our integral equation using some results of Ziman.[24] We restate our principle as a theorem. Let θ be the solution of the integral equation $I\theta = x$ and ψ be any function for which $(\psi,I\psi) = (\psi,x)$. Then of all functions ψ satisfying this relation, θ is the one which makes $(\psi,I\psi)$ a maximum. Thus, noting $\theta(s)$ is actually $B(\sigma,s)$

$$(\psi,I\theta) = \iint_{-\infty}^{\infty} \psi(f)\mathfrak{a}(f,s)\theta(s)dsdf$$
$$= \iint \psi(s)\mathfrak{a}(f,s)\theta(f)dsdf$$

from symmetry of $\mathfrak{a}(f,s)$. Adding gives

$$(\psi,I\theta) = \frac{1}{2}\iint \{\psi(f)\theta(s) + \psi(s)\theta(f)\}\mathfrak{a}(f,s)dsdf$$
$$= (\theta,I\psi) \tag{10}$$

from symmetry of integral between θ and ψ.

Putting $\theta = \psi$ we obtain integral with a square and if $\mathfrak{a}(f,s)$ is positive definite, we have for any ψ

$$(\psi,I\psi) \geq 0 . \tag{11}$$

Now since θ is the solution of the integral equation, the condition on ψ in the theorem implies

$$(\psi,I\theta) = (\psi,x) = (\psi,I\psi) .$$

From (11) we have

$$0 \leq ((\theta-\psi),I(\theta-\psi))$$
$$= (\theta,I\theta) + (\psi,I\psi) - (\theta,I\psi) - (\psi,I\theta)$$
$$= (\theta,I\theta) + (\psi,I\psi) - 2(\psi,I\theta) \quad \text{from (10)}$$
$$= (\theta,I\theta) - (\psi,I\psi) \quad \text{from (11)} ,$$

i.e.,

$$(\theta,I\theta) \geq (\psi,I\psi) \qquad \text{Q.E.D.}$$

Thus, we now write

$$(\theta,I\theta) = \iint_{-\infty}^{\infty} df ds B(\sigma,f)\mathfrak{a}(f,s)B(\sigma,s)$$
$$(\theta,x) = \int_{-\infty}^{\infty} df B(\sigma,f)\delta(f-\sigma) = B(\sigma,\sigma) .$$

[22] A. H. Wilson, *Theory of Metals* (Cambridge University Press, New York, 1953).

[23] M. Kohler, Z. Physik **124**, 772 (1948); **125**, 679 (1949).

[24] J. M. Ziman, Can. J. Phys. **34**, 1256 (1956).

To maximize the variational function using a Lagrange multiplier to include the constraint

$$\delta\{(\theta,I\theta) + \lambda[(\theta,I\theta) - (\theta,x)]\} = 0$$

$$\delta\{(1+\lambda)(\theta,I\theta) - \lambda(\theta,x)\} = 0$$

$$\delta\left\{(1+\lambda)\iint B(\sigma,f)\mathfrak{A}(f,s)B(\sigma,s)dfds - \lambda\int B(\sigma,f)\delta(f-\sigma)df\right\} = 0$$

$$\delta\left\{(1+\lambda)\iint B(\sigma,f)\mathfrak{A}(f,s)B(\sigma,s)dfds - \lambda B(\sigma,\sigma)\right\} = 0.$$

Divide by $1+\lambda$ and let $\mu = \lambda/(1+\lambda)$

$$\delta\left\{\iint B(\sigma,f)\mathfrak{A}(f,s)B(\sigma,s)dfds - \mu B(\sigma,\sigma)\right\} = 0\,.$$

Suppose we expand B in some suitable series of the form

$$B(f,s) = \sum_n \alpha_n\theta_n(f,s) \tag{12}$$

over a finite region $a \le f,\ s \le b$ where the θ_n are known functions making up a complete set and the coefficients α_n are to be chosen so that $B(f,s)$ is a solution of the original integral equation $I\theta = x$. We should then get integrals of the form $(\theta_n,I\theta_m)$ and (θ_n,x) that are coefficients of quadratic and linear forms in α_n. Thus, the variational principle is applied to get the best choice of the α_n by differentiation with respect to these variables (making the variational function a maximum subject to the subsidiary condition). The result is a set of linear equations to be solved for the α_n. For an exact solution we have an infinite sequence and therefore an infinite determinant to be evaluated. Careful choice of the θ_n may result in a good approximation with only a few terms and choice of the θ_n therefore depends on the insight which can be obtained in a given physical problem. We think at this point in terms of an exact solution to be evaluated to any desired degree of accuracy by the Ritz method.[25] Thus, in principle *each* B can be found from *each* $\mathfrak{A}$. If we now let $\mathfrak{A}$ and therefore B be random processes, the α_n become random variables to be evaluated.

In order to obtain a little more insight into this forbidding problem, we digress now to review some properties of Green's functions which we can use.

Consider then the following method[26] of finding a Green's function for ordinary nonstochastic physical problems.

Let L be a differential operator and s a continuous function. Consider the problem of finding the function f satisfying the inhomogeneous differential equation

$$Lf(x) = s(x) \tag{13}$$

and certain specified boundary conditions. If there exists a unique solution f for each s, there must exist an inverse operator L^{-1} such that for all s the formal solution will be

$$f(x) = L^{-1}s(x)\,. \tag{14}$$

The inverse operator implies not only an operation inverse to L but also application of the associated boundary conditions.

The solution of the differential equation (13) corresponding to $s(x) = \delta(x-x')$ is the Green's function $G(x,x')$ for the given L and boundary conditions. Thus, G satisfies

$$LG(x,x') = \delta(x-x')$$

with the same boundary conditions as on f. Thus,

$$G(x,x') = L^{-1}\delta(x-x') \tag{15}$$

and (14) can be rewritten

$$f(x) = \int L^{-1}\delta(x-x')s(x')dx'$$

or, using (15),

$$f(x) = \int G(x,x')s(x')dx'\,,$$

which is obvious anyway from the meaning of G or from applications, e.g., in electromagnetic theory.

A particularly nice method of obtaining the Green's function exists for self-adjoint (Hermitian) operators. Thus if $L = L^\dagger$, the eigenfunctions of L form a complete set. These eigenfunctions ψ_λ and the corresponding eigenvalues λ are found by solving the eigenvalue problem

$$L\psi_\lambda(x) = \lambda\psi_\lambda(x)$$

with the same boundary conditions on ψ_λ as on $f(x)$. We require also the closure condition

$$\int_\lambda \psi_\lambda(x)\psi_\lambda^*(x') = \delta(x-x')$$

where $\int_\lambda$ is used to mean $\int d\lambda$ if the spectrum of eigenvalues is continuous, or $\sum_\lambda$ if it is discrete. The Green's function is given by an expansion over the set of ψ thus,

$$G(x,x') = \int_\lambda \frac{\psi_\lambda(x)\psi_\lambda^*(x')}{\lambda}\,, \tag{16}$$

[25] F. B. Hildebrand, *Methods of Applied Mathematics* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1952).

[26] G. H. Goertzel and N. Tralli, *Some Mathematical Methods of Physics*, McGraw-Hill Book Company, Inc., New York, 1960).

so long as $\lambda \neq 0$ and no infinite sequence of λ's approach zero as a limit point.[27]

Now one method which suggests itself is to use such direct methods as the above for determining G by taking one member of the ensemble of the operator L, say L_i, determining a single G_i, and then averaging over the ensemble by using time averaging—assuming an ergodic property holds for the operator. This can be done when an explicit L_i is known and sometimes by considering a simpler nonrandom problem. [An example is considered in Chap. 4 of the referenced dissertation and in a paper (to be published).]

We ask ourselves now whether any expansion analogous to (16) exists for a stochastic operator. Analogous to the self-adjoint or Hermitian property for ordinary operators, does there exist a corresponding property for stochastic operators? We might say a stochastic operator is self-adjoint if it is *stochastically self-adjoint* meaning all members of the ensemble are self-adjoint, or we could consider a *self-adjoint-in-the-mean* property. Thus, suppose that the mean operator is represented by the matrix whose elements are the means of the elements of the stochastic matrix which corresponds to the stochastic operator. This mean operator might well be self-adjoint and possess an eigenfunction expansion. We can then speak of the operator as being self-adjoint-in-the-mean. Now hypothesize a stochastic expansion (for the stochastic operator) whose mean is the above mean eigenfunction expansion. The stochastic Green's function could then be represented in terms of the eigenfunctions and eigenvalues in a manner analogous to the deterministic case, i.e., in terms of the stochastic expansion. We have shown we can write the Green's function solution for a deterministic differential equation $Ly = x$ in terms of the eigenfunctions of L. Regard this L as a single representation of the stochastic operator $\mathfrak{L}$, i.e., one of the ensemble of operations represented by $\mathfrak{L}$. The eigenfunction expansion suitable for $\mathfrak{L}$ must be the eigenfunction expansion for L when $\mathfrak{L}$ becomes L, i.e., becomes nonrandom. Thus, if $\mathfrak{L}$ involves Gaussian parameters, in the limit as their variance decreases to Dirac δ-function behavior, we get L and the expansions now correspond. Imagine that all the representations of $\mathfrak{L}$ are expandable in the same basis—which can be true for a properly defined class of $\mathfrak{L}$. Then $\mathfrak{L}$ has the same set of eigenfunctions as L. Now, however, the coefficients of the expansion are random variables. (Only the eigenvalues are random. When these random variables become nonrandom we get the set for L.) If we write y in terms of x and a Green's function written from the above expansion (analogous to the determination of the Green's function for L), multiply by itself displaced by τ, take the mean, and transform to get $\Phi_y(f)$, we have a kernel or stochastic Green's function involving the mean of two Green's functions which involve random variables from the expansion for $\mathfrak{L}$. Taking the mean involves only the random coefficients of the expansion for $\mathfrak{L}$ and therefore the autocorrelation of those quantities which we will want to relate to known statistics of the operator. We define a stochastic operator to be self-adjoint when its every realization is self-adjoint. Such a stochastically self-adjoint operator is also necessarily self-adjoint in the mean.

Every realization or representative L_i of an ensemble $\{L_i\}$ representing a stochastic operator $\mathfrak{L}$ has the same basis if and only if $L_iL_j - L_jL_i = 0$ for all i and j, i.e., if every realization commutes with every other.

When the conditions of the last statement are not satisfied, i.e., when there does not exist a common basis, a stochastic Green's function or kernel K_H is formed in the same manner with the difference that the eigenfunctions as well as the eigenvalues are dependent on the random variable and the averaging process involves these as well. For stochastic *differential* operators, this is expected to be the usual case since only a very restricted class of self-adjoint operators would allow different realizations of $\mathfrak{L}$ to commute. For other operators than differential forms, there are more possibilities. This then appears to be an alternate method of expressing the kernel K_H and a more direct one since it gives K_H in terms of the statistics of $a_\nu(t)$ for problems in which Green's functions can be found. (This method will be discussed further in Sec. 5.) We also saw that K_H could be expressed in terms of $\langle B\, B^* \rangle$, which means $\langle B\, B^* \rangle$ can be written in terms of statistics of the a_ν and relations exist between these quantities which it is worthwhile to explore. Thus, in Eq. (13), writing $\mathfrak{L}$ to emphasize according to the above discussion the stochastic nature of the operator, each realization or member or sample function $\mathfrak{L}_i$ of the ensemble corresponding to $\mathfrak{L}$ has a Green's function $G_i(x,x')$. Correspondingly, each sample function of the output satisfies

$$y_i = \int G_i(x,x')s(x')dx' , \tag{17}$$

and the process y is the entire ensemble of the y_i, i.e., $\{y_i\}$. However, we are not solving for y but for an

[27] Reference 26, p. 167.

appropriate statistical measure and we have chosen power spectra as a convenient one, thus,

$$\Phi_y = \int ds\langle |B(s,f)|^2\rangle \Phi_x(s) \tag{18}$$

or, for convenience in writing,

$$\Phi_y = \int ds C(s,f)\Phi_x(s) \tag{19}$$

or alternatively a matrix product $C\cdot\Phi_x$. The C is the (stochastic) Green's function for the appropriate statistical measure (in this case, the spectral densities) of the functions instead of for the functions themselves. C is the same as K_H of course, except that H has no particular significance here. From $\{y_i\}$ one can obtain a corresponding power spectrum Φ_y. If the output is stationary, Φ_y could be obtained from (17) since each sample function or member then gives the same spectrum. We assumed in deriving Eq. (9) that the input is stationary. Thus, it is necessary that the stochastic operator have stationary properties, which is why we specified the coefficients could be translated without affecting the statistical properties under consideration. The C in (19) clearly corresponds to some mean combination of the G_i and is therefore expandable in the set ψ_λ appropriate to a member $\mathfrak{L}_i$ of the $\mathfrak{L}$ process but involves an appropriate averaging process. Thus, from (17) we can write

$$R_{y_i}(\tau) = \langle y_i(x)y_i^*(x+\tau)\rangle$$

and

$$\begin{aligned}\Phi_y(f) = \Phi_{y_i}(f) &= \int_{-\infty}^{\infty} e^{2\pi i f\tau} R_{y_i}(\tau)d\tau \\ &= \int_{-\infty}^{\infty} d\tau e^{2\pi i f\tau}\int_{-T/2}^{T/2} d\alpha \int_{-T/2}^{T/2} d\beta \\ &\quad\times \langle G_i(x,\alpha)G_i^*(x+\tau,\beta)\rangle\langle s(\alpha)s^*(\beta)\rangle\,.\end{aligned}$$

If the input is stationary, let $\beta = \alpha + \delta$

$$\langle s(\alpha)s^*(\alpha+\delta)\rangle = R(\delta)$$

$$\begin{aligned}\Phi_y(f) &= \int ds\int d\tau\int d\alpha\int d\delta e^{2\pi i(f\tau - s\delta)} \\ &\quad\times\langle G_i(x,\alpha)G_i^*(x+\tau,\alpha+\delta)\rangle\Phi_x(s) \\ &= \int ds C(s,f)\Phi_x(s)\end{aligned}$$

so that $C(s,f) = K_H(s,f) = \langle |B(s,f)|^2\rangle$ satisfies

$$\begin{aligned}C(s,f) &= \int d\tau\int d\alpha\int d\delta e^{2\pi i(f\tau - s\delta)} \\ &\quad\times\langle G_i(x,\alpha)G_i^*(x+\tau,\alpha+\delta)\rangle\,.\end{aligned}$$

By use of the ergodic hypothesis, we can replace the ensemble average by an integration over x which corresponds to time in this notation.

$$\begin{aligned}C(s,f) &= \int_{-\infty}^{\infty} d\tau\int_{-T/2}^{T/2} d\alpha\int_{-T/2-\alpha}^{T/2-\alpha} d\delta e^{2\pi i(f\tau - s\delta)} \\ &\quad\times\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} dx G_i(x,\alpha)G_i^*(x+\tau,\alpha+\delta)\,.\end{aligned} \tag{20}$$

Thus, for ergodic operators we could write the kernel C (or K_H) from knowledge of a solution for a deterministic member. We might then conceive of a future study of deterministic equations or systems by imagining them embedded in a larger class or ensemble and expressing functional relations among members of this class to obtain further insights than could be obtained through consideration of the original process alone, an idea also considered by Bellman and Kalaba.[28] For some problems (20) may be sufficient, however, for the problem at hand, the expressions obtained are not satisfactory since we want a solution in terms of the statistics of the coefficients (their covariance) rather than in terms of a representative or sample function of the operator.

Returning to the general discussion, we are now concerned with cases where the stochastic operator is not ergodic or does not otherwise lend itself to the Green's function expansion method. In such cases (as well as for the sake of discovering all relationships even if we can get the Green's functions in terms of the $\mathfrak{L}$ directly), we are once again interested in the B quantities defined earlier. In Eq. (12) the complete set θ_n was said to be arbitrary. However with our differential equation $Ly = x$ (we no longer emphasize the stochastic nature of L by writing $\mathfrak{L}$, since we need not distinguish it from a deterministic L), we, in general, have some boundary conditions on $y(t)$. If these are either given directly or can be translated to restrictions on $\Phi_y(f)$, these restrictions must also hold on the kernel K_H (which is the output for a particular input, the δ function). Therefore, it is possible that the expansion functions θ_n can be chosen to satisfy these restrictions and the α can then be determined by our variational principle. Then, $B(s,f)$ is evidently known in terms of $\mathfrak{A}(s,f)$, i.e., in terms of the coefficients of the differential equation [see Eqs. (6), (7), and (9)].

The desired kernel is

$$\begin{aligned}\langle |B(s,f)|^2\rangle &= \int d\sigma\delta(s-\sigma)\langle B(\sigma,f)B^*(s,f)\rangle \\ &= \int d\sigma\delta(s-\sigma)\langle\textstyle\sum_n \sigma_n x_n\theta_n(\sigma,f)\ \sum_m \sigma_m^* x_m^*\theta_m^*(s,f)\rangle\,,\end{aligned} \tag{21}$$

[28] R. Bellman, and R. Kalaba, *Transaction of the 1959*

where we have substituted the form (12) for $B(f,s)$ which will be taken to be a Karhunen–Loeve expansion.[29]

This is just a generalized Fourier series or orthogonal expansion such that the coefficients are uncorrelated. A nonperiodic random process cannot, of course, be written as a trigonometric Fourier series with uncorrelated random coefficients but it is possible to obtain uncorrelated coefficients by using a Karhunen expansion instead of a Fourier series. Thus, a random process, in this case $B(f,s)$, is expanded over a domain or region $a \leq f,s \leq b$ in the form $\sum_n \alpha_n \theta_n(f,s)$. Letting $\alpha_n = \sigma_n x_n$ where the σ_n are numbers, possibly complex, and the x_n are random variables, it follows that

$$\langle x_m x_n^* \rangle = \delta_{mn}$$

and

$$\int_a^b \theta_n \theta_m^* df ds = \delta_{mn} .$$

What we mean precisely by equating B with the series is that for every f,s in the domain, we have the limit-in-the-mean (l.i.m.).

$$B(f,s) = \text{l.i.m.}_{N\to\infty} \sum_{n=1}^{N} \sigma_n x_n \theta_n(f,s)$$

or

$$\lim_{N\to\infty} \langle |B(f,s) - \sum_{n=1}^{N} \sigma_n x_n \theta_n(f,s)|^2 \rangle = 0 .$$

Actually, any random process, even a nonstationary one, has an orthogonal expansion of this type (with uncorrelated and random coefficients). If the orthonormal set used as a basis were that of the ordinary or trigonometric Fourier series, the coefficients would be correlated unless the process were periodic. Suppose we can equate in the specified sense the process $B(f,s)$, with the given series for some set of random variables x_n.

Equation (21) becomes

$$\begin{aligned} \langle |B(s,f)|^2 \rangle &= \int d\sigma \delta(s-\sigma) \\ &\quad \times \langle \textstyle\sum_n \sigma_n x_n \theta_n(\sigma,f) \sum_m \sigma_m^* x_m^* \theta_m^*(s,f) \rangle \\ &= \int d\sigma \delta(s-\sigma) \textstyle\sum_n |\sigma_n|^2 \theta_n(\sigma,f) \theta_n^*(s,f) \\ &= \textstyle\sum_n |\sigma_n|^2 \theta_n(s,f) \theta_n^*(s,f) . \end{aligned} \tag{22}$$

Thus, once the set θ_n has been determined so that B could be obtained by the variational method, it becomes simple to get the kernel itself, i.e., $\langle |B(s,f)|^2 \rangle$ or $K_H(s,f)$.

However, it is not easy to find the appropriate set θ_n. What we would like is a prescription for finding the θ_n appropriate to the differential operator, boundary conditions, and the given distribution functions or other knowledge of the statistics of L. Let us examine what is involved. Multiplying both sides of (22) by $\theta_k^*(s,f)$, integrating over the domain $a \leq f,s \leq b$, and using the orthonormality property of the θ's, we get

$$\iint_a^b ds df K_H(s,f) \theta_k(s,f) = \lambda_k \theta_k(s,f) ,$$

where $\lambda_k = |\sigma_k|^2$.

Thus the θ_k are the eigenfunctions of the integral operator $\int\int_a^b ds\, df\, K_H(s,f)$ for eigenvalues λ_k, i.e., the solutions of a Fredholm integral equation. If solved, this equation would give us numbers σ_k and the set θ_n for our expansion of $B(f,s)$. We can assume σ_k to be nonzero and eigenvalues of multiplicity $r > 1$ to be indexed with r different numbers so a Schmidt orthogonalization procedure[30] can be employed. Take σ_k to be the positive square root of $|\sigma_k|^2$.)

However this solution requires knowledge of the B or at least of the kernel $\langle B\, B^* \rangle$ so the basis functions for B cannot be obtained without knowing B. In principle, we might consider assuming a basis θ_n possibly from knowledge of the boundary conditions, finding B then solving the Fredholm integral equation until we get the same basis. The difficulty in this procedure is that our variational method gives us a sample B for a sample $\mathfrak{A}$ but what we want are expressions in terms of means of $\mathfrak{A}$ or covariances of a_ν.

We return now to our condition defining B in terms of $\mathfrak{A}$, i.e.,

$$\int_{-\infty}^{\infty} ds \mathfrak{A}(f,s) B(\sigma,s) = \delta(f-\sigma) .$$

If we write $B(s,\sigma)$ instead of $B(\sigma,s)$—this was an arbitrary choice—and conceive the *function* $\mathfrak{A}(f,s)$ to be a matrix by imagining the f as representing a continuous row index and s a continuous column index, the left-hand side looks like a matrix product while the right-hand side is similar to the identity matrix except that the Dirac δ in place of a Kronecker δ implies ∞ on the diagonal rather than a series of 1's. By analogy we are justified in applying the term "inverse" to B. More specifically, we call B an *inverse*

International Symposium on Circuit and Information Theory, Los Angeles, June 16–18, 1959. IRE Trans. Circuit Theory, **CT-6** (Special Supplement) (1959).

[29] K. Karhunen, *Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung* (Helsinki, 1947). Also, M. Loeve, *Probability Theory* (D. Van Nostrand Company, Princeton, New Jersey, 1960).

[30] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953); F. Mandl, *Quantum Mechanics* (Academic Press Inc., New York, 1959).

functional of $\mathfrak{a}$. Now, it is obvious that $\langle \mathfrak{a}\mathfrak{a}^* \rangle$ can be given explicitly in terms of the correlations R_{a_ν} for the random coefficients $a_\nu(t)$ in $\mathfrak{L}$. To see this we use the definition of A_ν and write

$$\langle A_\nu(s-f)A_\nu^*(s-f)\rangle = \iint_{-T/2}^{T/2} \langle a_\nu(u)a_\nu(v)\rangle du dv .$$

Let $v = u + \tau$, then the above quantity becomes

$$\int_{-T/2}^{T/2} du \int_{-T/2-u}^{T/2-u} d\tau R_{a_\nu}(\tau) .$$

Finally,

$$\langle \mathfrak{a}(f,s)\mathfrak{a}^*(f,s)\rangle = \sum_\mu \sum_\nu (-1)^\mu (2\pi i f)^{\mu+\nu} \times \int_{-T/2}^{T/2} du \int_{-T/2-u}^{T/2-u} d\tau R_{a_\nu}(\tau) . \quad (23)$$

If we could write B in terms of $\mathfrak{a}$, form $\langle B\, B^* \rangle$ and possibly expand in a series of functions or powers of $\langle \mathfrak{a}\, \mathfrak{a}^* \rangle$, we might have a computational method of interest. Of course, we are still suggesting the inversion of a matrix, but a much simpler one than the problem of inversion of the matrix corresponding to $\mathfrak{L}$. Our matrices are continuous and infinite here but can be made discrete by originally using series instead of transforms. Consider the case of discrete, ordinary or deterministic matrices and the matrix product $AB = I$, where I is the identity matrix. Now B is called the inverse or reciprocal matrix and is indicated by A^{-1}. (It does not of course mean $1/A$ or even I/A since division by matrices is undefined). To see how A^{-1} is written in terms of A let A be represented by the array of elements:

$$A = \begin{Vmatrix} a_{11}\, a_{12} \cdots a_{1n} \\ a_{21}\, a_{22} \cdots \\ \cdot \\ \cdot \\ \cdot \\ a_{n1}\, a_{n2} \cdots a_{nn} \end{Vmatrix} .$$

We know, of course, that the determinant of A, or $|A|$, can be expanded by cofactors as

$$|A| = \sum_j a_{ij}A_{ij} = \sum_i a_{ij}A_{ij} ,$$

if we define A_{ij} as the cofactor of a_{ij}. We know further that $|A|$ vanishes if any two rows or columns of A are the same. Thus,

$$\sum_j a_{kj}A_{ij} = \sum_j a_{ji}A_{jk} = 0 \quad i \neq k .$$

Combining these statements

$$\sum_j a_{ij}A_{kj} = \sum_j a_{ji}A_{jk} = |A|\delta_{ki} , \quad (24)$$

which says simply that expansion by a row or column i in terms of the cofactors of a row or column k vanishes if $i \neq k$ and is equal to $|A|$ when $i = k$.

Now Eq (24) looks like a matrix product. Set $\hat{a}_{ij} = A_{ji}$ and (24) becomes

$$\sum_j a_{ij}\hat{a}_{jk} = \sum_j \hat{a}_{kj}a_{ji} = |A|\delta_{ki} ,$$

which can be written in matrix form as

$$A\hat{A} = \hat{A}A = |A|I$$

if

$$\hat{A} = \left\| \hat{a}_{ij} \right\| = \begin{Vmatrix} A_{11}\, A_{21} \cdots A_{n1} \\ A_{12} \cdots \\ \cdot \\ \cdot \\ A_{1n} \cdots \quad A_{nn} \end{Vmatrix} .$$

Now following Hadley[31] we define

$$A^{-1} = (1/|A|)\hat{A} . \quad (25)$$

Then

$$AA^{-1} = A^{-1}A = A\hat{A}/|A| = I .$$

If A is nonsingular, it has an inverse defined by (25) and further, the inverse is unique.

To sum up we say if $AB = I$ then $B = \hat{A}/|A|$ where we write $\hat{A}$ to mean the so-called adjoint matrix of A where this overworked term is not to be confused with the Hermitian adjoint $A^\dagger$ since $\hat{A}$ is found from the transpose of the cofactors of the elements of A not the transpose of the complex conjugates of the elements of A.

Thus, in our own problem with the defined (but discretized) $\mathfrak{a}$ and B we write

$$\langle BB^* \rangle = \left\langle \frac{\hat{\mathfrak{a}}\hat{\mathfrak{a}}^*}{|\mathfrak{a}||\mathfrak{a}|^*} \right\rangle = \left\langle \frac{\hat{\mathfrak{a}}\hat{\mathfrak{a}}^*}{|\mathfrak{a}\mathfrak{a}^\dagger|} \right\rangle ,$$

which means precisely the average of the following array

$$\frac{\begin{Vmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ A_{1n} & & & A_{nn} \end{Vmatrix} \begin{Vmatrix} A_{11}^* & A_{21}^* & \cdots & A_{n1}^* \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ A_{1n}^* & & & A_{nn}^* \end{Vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & \\ a_{21} & a_{22} & \cdots & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{n1} & & & \end{vmatrix} \begin{vmatrix} a_{11}^* & a_{12}^* & \cdots & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{n1}^* & & & a_{nn}^* \end{vmatrix}} ,$$

[31] G. Hadley, *Linear Algebra* (Addison–Wesley Publishing Company, Inc., Reading, Massachusetts, 1961).

where a_{ij} are the elements of $\mathfrak{A}$ and A_{ij} are the cofactors of a_{ij}.

The difficulty in finding $|\mathfrak{A}|$ and the above expression is avoided in quantum theory by using approximate representations in which $\mathfrak{A}$ is diagonal. The general problem here appears to be enormously difficult even with arbitrary limitations of the number of rows and columns for successive approximations. Since we do know $\langle \mathfrak{A}\, \mathfrak{A}^* \rangle$ explicitly it is logical to ask if we can write our expression in some series of powers of $\langle \mathfrak{A}\, \mathfrak{A}^* \rangle$ or some function of $\langle \mathfrak{A}\, \mathfrak{A}^* \rangle$ which could then be written as a function of R_{a_ν}. It is not clear how the desired expansion is to be made so the problem is not completed. In the referenced dissertation, a differential equation with a random Gaussian coefficient is studied at length and a solution is obtained. Perhaps then if the expansion is used for the same problem, we will be able to see how it can be done or if $\langle \mathfrak{A}\, \mathfrak{A}^* \rangle$ is already the first term. We know the series terminates in that problem and we can see that the kernel is a function of the correlation of the random coefficient alone which is easily calculable. It seems reasonable to expect that our desired calculation can be made for that same (Gaussian) case.[32]

We will make a few remarks about the existence of B given an $\mathfrak{A}$ and then go on to other methods.

3. Necessary and Sufficient Conditions for the Existence of B

Let W be the class of all $\mathfrak{L}_2$ integrable functions ($\mathfrak{L}_2$ is the vector space of all real-valued functions whose square is Lebesgue integrable over the interval). Then, as is well known, B exists and is unique if and only if

$$\int_{-\infty}^{\infty} \mathfrak{A}(f,s)W(f)df = 0$$

implies $W(f) = 0$, i.e., the integral operator $\int_{-\infty}^{\infty} \mathfrak{A}(f,s)df$ does not have zero in its spectrum.[33] (For discrete $\mathfrak{A}$ this may be the case and can be seen easily. For continuous $\mathfrak{A}$ it's very difficult to decide.)

If $\mathfrak{A}$ doesn't have this property, there exists no B satisfying our condition. As an example, take

$$\mathfrak{A}(f,s) = e^{-|f-s|}.$$

The inverse is an unbounded second-order differential operator with zero in its spectrum. No B exists.

[32] An iterative method which appears very promising is now being investigated and will be discussed in a forthcoming paper.

[33] See, e.g., M. H. Stone, Am. Math. Soc. Colloq. Publ. 15 (1937).

4. Karhunen–Loeve Expansions

The solution we seek is dependent on appropriate expansions of the random processes involved. Our goal is to evaluate $\langle B\, B^* \rangle$ in terms of knowledge of the statistics of the $a_\nu(t)$. If the a_ν are Gaussian processes and their cross correlations are zero, they are statistically independent. Knowledge of their autocorrelations R_{a_ν} is sufficient for complete statistical knowledge. Whatever the distribution functions involved however, the $a_\nu(t)$ can always be expanded in Karhunen–Loeve (K.L.) expansions. We consider briefly the nature of such expansions.

An expansion for a random process $x(t)$ on an interval (a,b) of the form $\sum_n \sigma_n x_n \phi_n(t)$, where the σ_n are numbers, the x_n are random variables satisfying $\langle x_m x_n^* \rangle = \delta_{mn}$, and the ϕ_n are orthonormal, i.e.,

$$\int_a^b \phi_n \phi_m^* dt = \delta_{mn}\,,$$

is a K. L. expansion of the random process on (a,b). What we mean precisely by equating $x(t)$ with this series is that for every t in (a,b) we have the limit-in-the-mean

$$x(t) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{n=1}^{N} \sigma_n x_n \phi_n(t)$$

or

$$\lim_{N\to\infty} \langle |x(t) - \sum_{n=1}^{N} \sigma_n x_n \phi_n(t)|^2 \rangle = 0\,.$$

Any random process, even a nonstationary one, has an orthogonal expansion of this type with random but uncorrelated coefficients. If the orthonormal set used as a basis were that of the ordinary trigonometric Fourier series, the coefficients would be correlated unless the process is periodic.

To find the appropriate set ϕ_n and the numbers σ_n we suppose we can equate, in the indicated sense, the process $x(t)$ with the given series for some set of random variables x_n.

$$\langle x(t)x^*(s) \rangle = R_x(t,s) = \langle \sum_n \sigma_n x_n \phi_n(t) \sum_k \sigma_k^* x_k^* \phi_k(s) \rangle$$
$$= \sum_n |\sigma_n|^2 \phi_n(t)\phi_n^*(s) \;;\; (a \le t \le b)\,,$$

where we have now defined the autocorrelation function $R_x(t,s)$. Multiplying $R(t,s)$ by $\phi_k(s)$ and integrating over s in (a,b) and using orthonormality of the ϕ_n

$$\int_a^b ds R(t,s)\phi_k(s) = |\sigma_k|^2 \phi_k(t)\,,$$

which we recognize as an eigenvalue equation for the integral operator $\int_a^b ds\, R(t,s)$ for eigenvalues λ_k

$= |\sigma_k|^2$, (or the solutions of a Fredholm integral equation). Solving this eigenvalue equation gives us the numbers σ_k and the orthonormal set ϕ_n for our expansion. Assume σ_k to be nonzero (and eigenvalues of multiplicity $r > 1$ to be indexed with r different numbers so a Schmidt orthogonalization procedure can be employed). Take σ_k to be the positive square root of $|\sigma_k|^2$. The random variables x_n are found in the usual manner for determining coefficients of expansions, thus,

$$\sigma_n x_n = \int_a^b x(t)\phi_n^*(t)dt .$$

Substituting, we quickly verify

$$\sigma_n\sigma_m\langle x_n x_m^*\rangle = \sigma_n\sigma_m\delta_{nm}$$

or

$$\langle x_n x_m^*\rangle = \delta_{nm}$$

as specified at the start. Further, denoting $\sum_{n=1}^{N} \sigma_n x_n \phi_n(t)$ by $x_N(t)$, we can calculate

$$\langle x(t)x_N^*(t)\rangle = \langle x^*(t)x_N(t)\rangle = \langle x_N(t)x_N^*(t)\rangle = \sum_{n=1}^{N} \sigma_n^2\phi_n\phi_n^*$$

and hence,

$$\langle |x(t) - x_N(t)|^2\rangle = R(t,t) - \sum_{n=1}^{N} \sigma_n^2\phi_n\phi_n^* ,$$

where the last term on the right side converges to $R(t,t)$ as $N \to \infty$ by Mercer's theorem[34] establishing the statistical convergence we specified.

We note in passing that the energy in a random signal $x(t)$ over time interval (a,b) must be the average of the energy in different members of the process $x(t)$ or

$$\left\langle \int_a^b dt|x(t)|^2 \right\rangle = \sum_{n=0}^{\infty} \sigma_n^2 .$$

The solution of the eigenvalue equation can be carried out when the corresponding spectral density is rational,[35] i.e., when the Fourier transform of the autocorrelation function $R(s,t)$ is a rational function.

$$R_x(s,t) = R_x(s - t) = R_x(\tau) = \int_{-\infty}^{\infty} e^{2\pi i f\tau}\Phi_x(f)df$$

Φ_x is the spectral density of $x(t)$ and is assumed positive and integrable [so $R(\tau)$ is an autocorrelation] and even [so $R(\tau)$ is real], and *either rational or approximated by a sum of rational functions.* Thus, let $2\pi i f = p$,

$$R(\tau) = \int_{-\infty}^{\infty} e^{p\tau}\Phi_x(f)df$$

$$\Phi_x(f) = \sum_k \Phi_x^{(k)}(f) = \frac{2}{\pi}\sum_{k=1}^{n} \frac{A_k c_k}{c_k^2 - p^2} = \frac{N(p^2)}{D(p^2)} ,$$

where N and D are polynomials of degree n and d, respectively, d being greater than n, and $D(p^2)$ having no real roots.

5. Evaluation of the Kernel

The example

$$dy/dt + \xi(t)y = \eta(t) ,$$

where ξ and η are random is considered at length in the referenced dissertation.[36] Suppose $a_\nu(t)$ or in this case $\xi(t)$ is expanded in a K. L. series. In general then $a_\nu(t) = \sum_n (\gamma_\nu)_n\phi_n(t)$ or in this case,

$$\xi(t) = \sum_n \gamma_n\phi_n(t) = \sum_n \sigma_n\xi_n\phi_n(t)$$

in $a \le t \le b$, where the ϕ_n are orthonormal, the random variables ξ_n are uncorrelated, and the σ_n are numbers. Again, the set ϕ_k is found from an eigenvalue equation involving $R(t,s) = \langle\xi(t)\xi^*(s)\rangle$. If $\xi(t)$ is stationary, $R(t,s) = R(t - s) = R(\tau) = \langle\xi(t)\xi^*(t - \tau\rangle$.

Consider now the use of the K. L. coefficients instead of the Fourier transforms in defining the $\mathfrak{A}$ and B functions. Thus, we would substitute into the differential equation the following:

$$x(t) = \sum \alpha_n\phi_n(t)$$
$$y(t) = \sum \beta_n\phi_n'(t)$$
$$a_\nu(t) = \sum (\gamma_\nu)_n\phi_n''(t)$$

$$\alpha_n = \int_a^b x(t)\phi_n^*(t)dt = \sigma_n x_n = \text{K.L. transform of } x$$

$$\beta_n = \int_a^b y(t)\phi_n^{*\prime}(t)dt = \delta_n y_n = \text{K.L. transform of } y$$

$$(\gamma_\nu)_n = \int_a^b a_\nu(t)\phi_n^{*\prime\prime}(t)dt = \epsilon_n(a_\nu)_n = \text{K.L. transform of } a_\nu .$$

---

[34] Mercer's theorem for a positive-definite autocorrelation function $R(s,t)$ states

$$R(s,t) = \sum_{k=1}^{\infty} \lambda_k\phi_k(s)\phi_k^*(t) ,$$

where the convergence is uniform for s and t in the closed interval a,b (denoted $[a,b]$).

[35] The solution is made by Davenport and Root on p. 376. (See reference 7.)

[36] Also considered are generalizations to random vector fields and the wave propagation problem for a random inhomogeneous medium and an illustration of the generation of an approximate stochastic Green's function for such a problem from the simpler Green's function of a nonrandom case.

Thus we need to differentiate ϕ_n' ν times, multiply by ϕ_n^*, and integrate from a to b. The basis ϕ_n for x can be determined from R_x, the basis ϕ_n'' for $a_\nu(t)$ from R_{a_ν}; however, the basis ϕ_n' is unknown since it depends on the solution we seek, R_y, so we cannot hope to find the β_n unless we have some reason to choose the ϕ_n'.

We therefore expand only the coefficients $a_\nu(t)$ in K. L. series, each in its appropriate basis. We now have explicit, i.e., known and deterministic, time dependence in the $\mathfrak{L}$ and as many random variables as there were random coefficients in $\mathfrak{L}$. We now seek the eigenfunctions or basis functions which diagonalize the operator $\mathfrak{L}$. Treating the random variables as constants we solve $\mathfrak{L}\psi_\lambda = \lambda\psi_\lambda$. For all equations in which this can be carried out we can complete our problem since a Green's function can be written in terms of the ψ and λ. We limit ourselves to self-adjoint operators for $\mathfrak{L}$. Most physical problems can be considered to result from variational statements and the appropriate operations are always self-adjoint, i.e., we might consider the Hermitian property as being a physical realizability condition. Thus $\mathfrak{L}$ might conveniently be chosen to be of the general form

$$\mathfrak{L} = (d/dt)^n T_n (d/dt)^n ,$$

which is self-adjoint regardless of the distribution of T_n. If T_n is constant we get only even ordered derivatives. If $\mathfrak{L}$ is stochastic T_n is a function of time for which we have a distribution function. A second order operator $\mathfrak{L}$ acting on u is self-adjoint if the form of $\mathfrak{L}u$ is $fu'' + gu' + hu$ where h, g, f are functions of x and $g = f'$. We can take over Sturm–Liouville theory or in fact any method of ordinarily arriving at Green's functions.

Now, however, the Green's function depends on the random variable or variables. When we form $\langle y(t)y(t-\tau)\rangle$ then we will involve averaging over the random variables appearing in the average of two Green's functions. Where this average can be carried out, it gives us the kernel K_H or $\langle B\,B^*\rangle$. Thus in the first order equation with random $\xi(t)$ which is stationary and gaussian, the Green's function $h(t,u)$ will depend on the random variable ξ_n. The ξ_n has a joint Gaussian distribution so the averaging can be carried out. If it does turn out that all the members of $\mathfrak{L}$ can be expanded in a common basis (if they commute) we have $\mathfrak{L}\phi = \lambda\phi$ and the matrix elements $\mathfrak{L}_{nn} = \sum_n \phi_n^* \lambda_n \phi_n$ with only the λ being random, not the ϕ_n. The Green's function

$$G(x,x') = \sum_n \frac{\psi_n^{\xi_n}(x)\psi_n^{\xi_n}(x')}{\lambda_{\xi_n}}$$

would then have only the λ dependent on ξ_n and the averaging over ξ_n would be much simpler. This is the case in a restricted class of operators as we have discussed earlier.

The eigenvalue equation is in some respects easier than the differential equation although at first sight it appears equivalent. When the operator $\mathfrak{L}$ is stochastic, we actually have a non denumerable ensemble of differential equations which we must solve. An eigenvalue problem is a problem in matrix algebra which can be handled on digital computers. One can use finite difference methods to discretize a differential operator and get approximate answers with upper and lower bounds. If we have the matrix in the given problem, considerable labor is avoided.

To sum up: A Green's function for a differential operator L is the kernel of the integral operator that inverts L. However, we are considering kernels for statistical measures of the dependent variable y in $Ly = x$ rather than for y itself. For example, K_H is the kernel for Φ_y rather than for y. It corresponds to the kernel for the transform then of the mean of the product of $y(t)$ and $y^*(t+\tau)$. It's clear that the form of K_H is then just such a triple integral as found earlier. Each Green's function for y is a matrix; each observation or realization gives us a sample or representative matrix. The multiplication of $y(t)$ by $y^*(t-\tau)$ gives us a product of matrices. We call this product *Green's matrix*. When L becomes stochastic, the Green's matrix becomes stochastic. Since the correlation of y involved an ensemble average over a product, we have the average of the Green's matrix which is our stochastic Green's function. Each of the original Green's functions means one integration. The transform to power spectra gave the third integration. This triple integral involving an exponential (from the transformation) and an ensemble average over the matrix product is precisely our form for $K_H(s,f)$.

Thus we can take each sample or realization or representative of the stochastic operator, solve the eigenvalue equation and get one of the Green's functions discussed above. In the equation $dy/dt + \xi(t)y = \eta(t)$ we can solve for y as if ξ is not random and then carry out the statistical measure operation. To do it as stated we need to solve

$$[d/dt + \xi(t)]\phi = \lambda\phi$$

for a nonrandom ξ and write the Green's function in terms of the eigenfunctions ϕ.

The difficulty in doing this is that the explicit time dependence of even a representative ξ was unknown. When we made a K. L. expansion of $\xi(t)$

we obtained an explicit time dependence which we could handle.[37]

ACKNOWLEDGMENTS

I am very grateful to my advisor Professor David S. Saxon, to Professor A. V. Balakrishnan, Professor George E. Uhlenbeck, Professor Mark Kac, Professor Norbert Wiener, and the Physics Department staff and would like to express deep appreciation for their time and interest to them and to several others, particularly, Robert C. Heimiller, Dr. Donald F. DuBois, Dr. Dominic G. B. Edelen, and Dr. A. V. Gafarian.

Also, I would like to thank Dr. E. U. Condon for his editing and the management of Hughes Aircraft Company for substantial financial support and typing.

[37] Some of the examples of the dissertation and further applications will be discussed in forthcoming papers and in *Mathematics in Science and Engineering*, edited by Richard Bellman [Academic Press Inc., New York (to be published, 1963)].

REVIEWS OF MODERN PHYSICS VOLUME 35, NUMBER 1 JANUARY 1963

The Theory of Waves in Stratified Fluids Including the Effects of Gravity and Rotation*†

IVAN TOLSTOY

Columbia University, Hudson Laboratories, Dobbs Ferry, New York

CONTENTS

1. INTRODUCTION

THIS article outlines the linear theory of waves in stratified, compressible fluids in a gravity field, at rest with respect to inertial or rotating coordinates, and neglecting viscosity and heat conduction. From a mathematical standpoint, this is one of the simpler branches of hydrodynamics, since it involves the solution of second-order differential equations which are linear. Despite its formal simplicity, the field has not been truly preempted. Eckart's recent and interesting book[1] on the subject is clear testimony to this.

There are four principal effects in the assumed model: compressibility, stratification, gravity, and rotation. Considering individually the cases for which these features are either present or absent, one could construct many different models, the simplest of which, obtained by ignoring all but one or two of these effects, are well known and understood. Thus, one may think of sound waves, surface gravity waves, tidal oscillations, and the effect of Coriolis forces upon these. On the other hand, the effects of density stratification and gravity upon very-low-frequency sound in the atmosphere and the behavior of internal gravity waves in the oceans and atmosphere have certainly not been as systematically studied, although Eckart's work[1,2] has done much to remedy this situation. The formal simplicity of the equations may lead one to feel that the answers to all these problems, even if not actually known in detail, can at least be deduced from available solutions. This point of view, although justifiable perhaps from a purely mathematical standpoint, is of little actual help to the physicist involved with these questions. More often than not the propagation of the various possible modes of motion is dispersive and anisotropic, depending upon several physical parameters, and a surprising variety of possibilities lurks beneath the deceptively simple appearance of the equations.

* Columbia University, Hudson Laboratories Contribution No. 157.
† This work was supported by the Office of Naval Research.

[1] Carl Eckart, *Hydrodynamics of Oceans and Atmospheres* (Pergamon Press, New York, 1960).

[2] C. Eckart and H. G. Ferris, Rev. Mod. Phys. **28**, 48 (1956).