

WIENER'S CONTRIBUTIONS TO GENERALIZED HARMONIC ANALYSIS, PREDICTION THEORY AND FILTER THEORY

BY P. MASANI

0. Prologue. The strong cohesive forces permeating the work of Norbert Wiener complicate the task of surveying his contributions to specific areas. Where is one to begin and where to end? In the realm of prediction, for instance, Wiener's book [TS]¹ stands out as his first major contribution. But an important part of this book concerns the synthesis of predictors, for which as Kakutani remarked (32): "The theory of generalized harmonic analysis developed by the author some 20 years ago is exactly the right tool . . ." Now the latter theory, given in the memoir [GHA] of 1930, was itself the culmination of researches begun in 1924, which were motivated by even earlier investigations in the theory of Brownian motion. It would seem that a thorough review of Wiener's work in prediction should start from about the year 1919 when he looked at the Charles River from his office at M.I.T. and began to wonder whether the Lebesgue integral was the right tool for the analysis of the undulating water surface. Such a review would be beyond the abilities of this writer, even if he were granted the necessary space.

In this review we shall first survey those aspects of Wiener's great memoir [GHA] which bear on his later work on prediction and filtering (I). We shall then describe briefly how the mathematical activity of the thirties influenced his thought (II). Next we shall discuss Wiener's general theory of nonlinear prediction (III). From this we shall turn to his many contributions to linear prediction and filtering theory (IV). Lastly we shall dwell on his theory of filters (V).

I. GENERALIZED HARMONIC ANALYSIS (1930)

1. White light: the need for generalizing harmonic analysis. The optical origins of Wiener's work are best expounded from the standpoint of the electromagnetic theory of light. At a fixed point r in a medium traversed by light the direction of propagation at instant t is given by the Poynting vector $P(t) = E(t) \times H(t)$, where $E(t)$, $H(t)$

¹ The bold-faced numbers in brackets refer to the numbered references in the Bibliography of Norbert Wiener. The bold-faced letters in brackets and numbers in parentheses refer to the References at the end of this article.

are the electric and magnetic vectors at \mathbf{r} at instant t . The *optical vector* is identified with $\mathbf{E}(t)$; hence the instantaneous plane of optical vibration is that spanned by $\mathbf{E}(t)$ and $\mathbf{P}(t)$. When the light rays in the vicinity of \mathbf{r} have a direction \mathbf{z} independent of t , we can prescribe $\mathbf{E}(t)$ by its two components $E_x(t)$, $E_y(t)$ in (fixed) perpendicular directions \mathbf{x} , \mathbf{y} orthogonal to \mathbf{z} . The light signal is called *monochromatic* if E_x , E_y are *sinusoidal functions* of the same frequency;² otherwise *polychromatic*. Newton's experiments on dispersion revealed the polychromatic nature of sunlight, then called *white light*.

What sort of harmonic analysis can we make of a white light signal f ($= E_x$, say)?³ Classical theory offers but two alternatives:

(1) Assume that f is a generalized trigonometric polynomial or its limit:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{i\omega_k t}, \quad t \in (-\infty, \infty).$$

This assumption entails that the spectrum is made up of a finite or countable number of sharp lines of intensities $|c_k|^2$ at the frequencies ω_k , k = an integer. What is observed, however, is a continuous band of colors.

(2) Assume that $f \in L_2(-\infty, \infty)$. Then by the Fourier-Plancherel Theorem

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} c(\omega) e^{i\omega t} d\omega \quad (\text{in the l.i.m. sense}).$$

We now have as desired a continuous spectrum of energy on the frequency-band $(-\infty, \infty)$. But now

$$\int_t^{t+h} |f(\tau)|^2 d\tau \rightarrow 0, \quad \text{as } t \rightarrow \pm \infty \quad (h \text{ fixed}).$$

This means that the energy⁴ emitted by the signal during a time-interval of fixed length h approaches 0 as the interval advances (or recedes) indefinitely on the time axis. This result conflicts with the

² The function f is called *sinusoidal*, if $f(t) = A \cos(\omega t + \alpha) = \delta e^{-i\omega t} + ce^{i\omega t}$, $t \in (-\infty, \infty)$.

³ ω is the (*circular*) *frequency*, that is angular velocity, and $2\pi/\omega$ is the *period*. If $g(t) = ce^{i\omega t}$, $t \in (-\infty, \infty)$, then g is often called a *complex sinusoidal function of frequency* ω .

⁴ It will suffice to study the analysis of just one of the component functions E_x , E_y .

⁴ From Maxwell's equations the flux of electromagnetic energy through a small surface $(\delta S)\mathbf{r}$ is proportional to $|\mathbf{E}(t)|^2$, cf. (64, p. 333, (7-62)). Thus apart from a constant factor, $|E_\alpha(t)|^2$ represents the energy of the α -component.

standard theory in which the sun is treated as a permanent reservoir of infinite energy.⁵

Neither assumption seems to explain the behavior of white light satisfactorily. Wiener felt that the difficulty stemmed from the limitations of the classical harmonic analysis. Around 1924 he began developing a "generalized" analysis that could cover signals f on $(-\infty, \infty)$, which are on the one hand so irregular that their spectra are not made up of sharp lines alone, and on the other so lastingly vigorous that $\int_t^{t+h} |f(\tau)|^2 d\tau \rightarrow 0$, as $t \rightarrow \infty$. In this venture he benefited from the ideationally deep though logically unrigorous investigations of the physicists Rayleigh, A. Schuster and G. I. Taylor, and from the work of mathematicians such as Hahn, Bohr and Bochner.⁶ His researches culminated in his great memoir [GHA] of 1930. In this review we shall dwell only on those parts of [GHA] which bear on Wiener's later work in filter and prediction theory.⁷

NOTATION. We shall not adhere strictly to the notations and conventions used in [GHA] but adopt a few used by Wiener in later writings. For instance, we shall insert the factor $1/\sqrt{2\pi}$ in taking Fourier and Fourier-Stieltjes transforms, and write

$$(1.1) \quad \hat{f}(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} f(y) e^{-iyx} dy, \quad \tilde{f}(y) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} f(x) e^{ixy} dx.$$

We shall also use the abbreviations "SP" for "stochastic process," and BMSP for "the Brownian motion stochastic process."

2. The Wiener classes \mathcal{S} and \mathcal{S}' . As Wiener emphasized, the germs of generalized harmonic analysis were already in the work of Schuster, but a "radical recasting" of the latter was necessary [GHA, p. 118]. From this recasting emerged the *covariance function* ϕ of the signal f :

$$(2.1) \quad \phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t + \tau) \tilde{f}(t) dt, \quad \tau \in (-\infty, \infty).$$

Accordingly Wiener considered the class \mathcal{S} of all Borel measurable functions f on $(-\infty, \infty)$, for which ϕ is defined on $(-\infty, \infty)$, as well as the subclass \mathcal{S}' of $f \in \mathcal{S}$ for which ϕ is continuous at 0, and thence on $(-\infty, \infty)$.⁸ We can show that \mathcal{S} is closed under the trans-

⁵ This conception of the sun is of course untenable from a cosmological viewpoint, but it is a legitimate idealization for the study of short term optical phenomena.

⁶ The limitations of classical harmonic analysis had been felt in pure mathematics as well, for example in Dirichlet series, and in the statistical analysis of time series. The reader should consult the bibliography in [GHA] for references to earlier work.

⁷ Other aspects of [GHA] are perhaps discussed elsewhere in this number.

⁸ The classes $\mathcal{S}, \mathcal{S}'$ were actually introduced somewhat later, e.g., in [FI].

lation group (U_t , $-\infty < t < \infty$), $(U_t f)(\tau) = f(t + \tau)$, and that f and $U_t f$ have the same ϕ , [FI, p. 155]. But \mathcal{S} is not a vector space. Also, as is easy to check, ϕ is a positive definite function.

Wiener then introduced the function S of which ϕ is the Fourier-Stieltjes transform, calling it the *integrated spectrum* of f . (The current term is *spectral distribution* of f .) To appraise this work let us ignore order of precedence and proceed logically. Two years after the appearance of [GHA] came Bochner's extension to the real line of Herglotz's theorem on the Fourier-Stieltjes representability of continuous positive definite functions (5, §20). In 1939 Cramer (11) showed that the continuity assumption is dispensable. Now let $f \in \mathcal{S}$. Then since ϕ is positive definite, it follows from the Bochner-Cramer Theorem that

$$(2.2) \quad \phi(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-itu} dS(u), \quad \text{a.e.,}$$

where S is a bounded monotone increasing function, unique after proper normalization. It is easy to check that an admissible version of S is

$$(2.3) \quad S(u) = S(0) + \lim_{A \rightarrow \infty} \frac{1}{\sqrt{(2\pi)}} \int_{-A}^A \phi(t) \frac{e^{iut} - 1}{it} dt.$$

Wiener was obliged to proceed in a different way. He defined S by

$$(2.4) \quad \begin{aligned} S(u) = \text{l.i.m.}_{A \rightarrow \infty} \frac{1}{\sqrt{(2\pi)}} & \left(\int_{-A}^{-1} + \int_1^A \right) \frac{\phi(t) e^{iut}}{it} dt \\ & + \frac{1}{\sqrt{(2\pi)}} \int_{-1}^1 \phi(t) \frac{e^{iut} - 1}{it} dt, \end{aligned}$$

[GHA, (3.19)–(3.21)]⁹ and showed by hard work that it could be redefined on a set of zero measure so as to make it monotonic increasing and in fact to satisfy (2.3), [GHA, (3.27), (3.28)]. He then showed that S is bounded; and indeed that

$$(2.5) \quad \begin{aligned} \frac{1}{\sqrt{(2\pi)}} \{S(\infty) - S(-\infty)\} &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} \phi(t) dt \\ &\leq \phi(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |f(t)|^2 dt, \end{aligned}$$

⁹ The factor $1/2\pi$ occurs in [GHA] in place of $1/\sqrt{(2\pi)}$. But in later works, e.g., [FI, TS] Wiener uses $1/\sqrt{(2\pi)}$. Our choice of e^{iut} (and not e^{-iut}) in (2.4) and (2.3) agrees with that made in [GHA] and for SP's in [185, 186]. It has the merit of yielding holomorphic functions in the upper (and not lower) half plane in cases of interest, but the disadvantage that the power distribution of f is not $S/\sqrt{(2\pi)}$ but rather its “reflection” about the origin, cf. (4.4) et seq. below.

and that equality prevails if and only if $f \in \mathcal{S}'$ [GHA, (4.09), (4.10), (5.46)]. Only then was Wiener able to recover ϕ from S by the equation (2.2), [GHA, (5.40)].

We see that Wiener's treatment of S would have been appreciably simpler had the Bochner-Cramer results been at his disposal. But even in later writings he did not appeal to Bochner's theorem to define S but continued to use (2.4), cf. e.g., [TS, p. 42]. He did this perhaps to keep in sight the formal similarity between S and the important function s (below) which he had to define as a l.i.m.

The Wiener theory of S received an important generalization when Bochner defined S for the wider class of functions f such that

$$(2.6) \quad \lim_{T \rightarrow \infty} \frac{1}{\rho(T)} \int_{-T}^T f(t + \tau) \bar{f}(\tau) dt$$

exists for all t in $(-\infty, \infty)$, where ρ is any positive increasing function on $(0, \infty)$ for which $\rho(T+1)/\rho(T) \rightarrow 1$, as $T \rightarrow \infty$, (5, pp. 294–295).

We must next turn to another important function which Wiener associates with a function f in \mathcal{S} , viz. its *integrated* (or *generalized*) Fourier transform s . On comparing the \mathcal{S} -theory with the classical L_2 -theory we find that the covariance function ϕ is the natural counterpart of the inner product function ϕ^* in L_2 -theory defined by

$$\phi^*(t) = \int_{-\infty}^{\infty} f(t + \tau) \bar{f}(\tau) dt = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-itu} |s^*(u)|^2 du$$

where $s^* = \bar{f}$, (cf. (1.1)). Since this can be rewritten

$$\phi^*(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-itu} dS^*(u), \text{ where } S^*(u) = \int_{-\infty}^u |s^*(\lambda)|^2 d\lambda,$$

we see that Wiener's S is the counterpart of S^* . Does the Fourier-Plancherel transform s^* itself have a counterpart s in the \mathcal{S} -theory? If so, s would give the “generalized harmonic analysis” of an f in \mathcal{S} in somewhat the same way that s^* , i.e. \bar{f} , gives the “harmonic analysis” of an f in L_2 . Wiener gave an affirmative answer to this question and established for s generalizations of classical results for \bar{f} . In doing so he has left us with some hard analysis for which no simpler substitute seems to have been found during the last 35 years.

Wiener's work on s resembles to some extent that of Kolmogorov and Cramer on the spectral representation of a weakly stationary SP, a topic viewed best from the standpoint of Stone's spectral resolution for unitary flows on Hilbert spaces. The spectral representation involves stochastic integrals with values in the Hilbert space $L_2(\Omega, \mathfrak{B}, P)$. One might attempt to retrieve Wiener's results on s from this

representation by plugging in the random parameter ω . But many of these results pertain to the time-averages of (very irregular) functions in \mathcal{S} , and it is doubtful if any except the very trivial can be had in this way, cf. Doob (19).

We have thus to revert to Wiener's hard analysis involving "certain weighted averages" the appropriate tool for handling which is the "general theory of the Tauberian theorems developed by the author and applied to these problems by Mr. S. B. Littauer" [GHA, p. 119]. In this review we cannot go into this question. It will suffice to say that Wiener defines s by the equation obtained from (2.4) by replacing ϕ by f , and to mention two very interesting results concerning s . First,

$$(2.7) \quad \lim_{\epsilon \rightarrow 0} \frac{1}{4\pi\epsilon} \int_{-\infty}^{\infty} |s(u + \epsilon) - s(u - \epsilon)|^2 dx = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |f(t)|^2 dt;$$

in words, the "quadratic variation" of s equals the "mean-square modulus" of f [GHA, (5.53) & p. 119], [TS, (20.13)]. Next, [TS, p. 160, Theorem 28] an $f \in \mathcal{S}$ will be in \mathcal{S}' if and only if

$$(2.8) \quad \lim_{A \rightarrow 0} \limsup_{\epsilon \rightarrow 0} \frac{1}{4\pi\epsilon} \left(\int_{-\infty}^{-A} + \int_A^{\infty} \right) |s(u + \epsilon) - s(u - \epsilon)|^2 du = 0.$$

We should also point out the limitations of s . Whereas the spectral representation involves a stochastic integral, Wiener's corresponding equation [GHA, (6.01)]:

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-itu} ds(u), \quad \text{a.e.,}$$

makes sense only when the right-hand side is interpreted as the Cesaro limit of a pseudo-Stieltjes integral, i.e., one defined by a formal integration by parts. This is so because s is not of bounded variation except in trivial cases, [GHA, (6.02), (6.05)]. Let us not forget, however, that the *stochastic integral* which Cramer and others have used with much success, but which Wiener could not use in his \mathcal{S} -theory because of the absence of a random parameter, first appeared in Wiener's own work [29] of 1923, the integrator process of orthogonal increments being the BMSP. Let us add that Wiener also defined s for a wider class of functions than \mathcal{S} , viz. the set of f for which $(1/2T) \int_{-T}^T |f(t)|^2 dt$ is bounded, and the subset of the Bochner class for which the function ρ in (2.6) is $\rho(T) = T^n$, n being an odd positive integer [GHA, §6].

3. Multiple harmonic analysis and coherency. Wiener was able to

extend the foregoing ideas to vectorial signals with two or more components $f_1, \dots, f_q \in \mathcal{S}$, [GHA, §9]. We assume that for $1 < i, j < q$,

$$(3.1) \quad \phi_{ij}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f_i(t + \tau) \overline{f_j(t)} dt \quad \text{exists.}$$

Then by the Schwarz inequality

$$|\phi_{ij}(\tau)|^2 \leq \phi_{ii}(0)\phi_{jj}(0).$$

We are thus led to the $q \times q$ matrix-valued function $\phi = [\phi_{ij}]$, called the *covariance function* of the q -ple signal $\mathbf{f} = (f_1, \dots, f_q)$. ϕ_{ij} is called the *covariance* of f_i and f_j .

Again let us discard the historical order and proceed logically. The equations (3.1) reduce to the single matrix equation

$$(3.2) \quad \phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mathbf{f}(t + \tau) \mathbf{f}^*(t) dt$$

in which \mathbf{f} is a column vector and \mathbf{f}^* is its adjoint and therefore a row vector. The integrand is thus the $q \times q$ matrix with

$$f_i(t + \tau) \overline{f_j(t)}$$

in the (i, j) th entry. From this it readily follows that ϕ is a *function of non-negative type*, i.e.,

$$\phi(-t) = \phi(t)^*,$$

and for all $q \times q$ matrices $\mathbf{C}_1, \dots, \mathbf{C}_n$ and all real numbers t_1, \dots, t_n ,

$$\sum_{i=1}^n \sum_{j=1}^n \mathbf{C}_i \phi(t_i - t_j) \mathbf{C}_j^* = \text{a non-negative hermitian matrix.}$$

Hence by the matricial extension of Bochner's Theorem, cf. Cramer (11), (12)¹⁰ there exists a bounded $q \times q$ hermitian matrix-valued function $\mathbf{S} = [S_{ij}]$ on $(-\infty, \infty)$ with non-negative increments such that

$$(3.3) \quad \phi(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-it\lambda} d\mathbf{S}(\lambda), \text{ a.e.}$$

i.e.,

$$\phi_{jk}(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-it\lambda} dS_{jk}(\lambda), \text{ a.e.,} \quad 1 \leq j, k \leq q.$$

¹⁰ Our concept of *matricial non-negative type functions* is not delineated in Cramer's papers [11], [12]. But by introducing this concept we can get the matricial extension of Bochner's Theorem by using the very arguments which Cramer advances

Since each S_{jk} is of bounded variation and S_{jj} is real and monotonic increasing, these equations make sense. Again Wiener was obliged to proceed differently. He defined S_{jk} by

$$S_{jk}(u) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \phi_{jk}(t) \frac{e^{iut} - 1}{it} dt,$$

asserting that the last integral converges and that the matrix $[S_{ij}]$ so obtained has the properties just mentioned. (There is unfortunately some hand waving in this section of [GHA].)

The matrix function \mathbf{S} is called nowadays the *spectral distribution function* of the vectorial signal \mathbf{f} . Wiener called it the *coherency matrix* of \mathbf{f} . It is trivial to check that if $\mathbf{g}(t) = \mathbf{Af}(t)$, where \mathbf{A} is a constant $p \times q$ matrix, then the covariance and spectral matrices of the p -ple signal \mathbf{g} are $\mathbf{A}\Phi\mathbf{A}^*$ and $\mathbf{A}\mathbf{S}\mathbf{A}^*$. In particular taking $p=1$ we get the covariance and coherency matrices of the signal $f = \sum_{j=1}^q a_j f_j$:

$$(3.4) \quad \phi(t) = \sum_{j=1}^q \sum_{k=1}^q a_j \bar{a}_k \phi_{jk}(t), \quad S(\lambda) = \sum_{j=1}^q \sum_{k=1}^q a_j \bar{a}_k S_{jk}(\lambda).$$

Wiener defined two signals $f_1, f_2 \in \mathcal{S}$ to be *incoherent*¹¹ when the cross-covariance function ϕ_{12} or (equivalently) the cross spectral distribution S_{12} vanishes identically, i.e., when the covariance and spectral matrices of the 2-ple signal $\mathbf{f} = (f_1, f_2)$ are diagonal:

$$\Phi = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix}.$$

When the signals f_1, \dots, f_q are pairwise incoherent, (3.4) reduces to

$$\phi(t) = \sum_{j=1}^q |a_j|^2 \phi_{jj}(t), \quad S(\lambda) = \sum_{j=1}^q |a_j|^2 S_{jj}(\lambda).$$

4. Optical power, coherence and polarization. Wiener's theory of the \mathcal{S} class helped to clarify the optical concepts of power (i.e. intensity or brightness), coherence, and polarization. The statistical character of these concepts had been discerned by the physicists but they were unable to unravel the mathematical intricacies. The following account is based on Wiener's writings in optics [66], [168] and not just on [GHA] where the relevant exposition is somewhat sketchy.

(i) In view of the proportionality of $|\mathbf{E}(\mathbf{r}, t)|^2$ to the flux of electromagnetic energy through a small surface at \mathbf{r} perpendicular to the direction of propagation (cf. Footnote 4), it is natural to define

¹¹ He did not do so explicitly in [GHA], but his remarks suggest that he had it in mind, cf. [TS, p. 45].

$$(4.1) \quad \text{the total energy of } f = \int_{-\infty}^{\infty} |f(t)|^2 dt,$$

$$(4.2) \quad \text{the total power}^{12} \text{ of } f = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |f(t)|^2 dt = \phi(0).$$

Thus all signals f in \mathcal{S} have finite power and the nontrivial ones (for which $\phi(0) \neq 0$), such as steady white light signals, have nonzero power and infinite energy. Signals in $L_2(-\infty, \infty)$, so-called *transients* or *pulses*, have zero power and finite energy.

For the complex sinusoidal signal, $f(t) = ce^{i\omega t}$, $t \in (-\infty, \infty)$, it follows at once from (4.2) that the total power is $|c|^2$. Next for the signal

$$(4.3) \quad f(t) = \sum_{k=1}^n c_k e^{i\omega_k t}, \quad t \in (-\infty, \infty)$$

we easily find that

$$(4.4) \quad \phi(t) = \sum_{k=1}^n |c_k|^2 e^{i\omega_k t}, \quad S(u)/\sqrt{(2\pi)} = \sum_{-\infty < -\omega_k \leq u} |c_k|^2.$$

Thus $\{S(u_2) - S(u_1)\}/\sqrt{(2\pi)}$ is the power due to the frequencies in the interval $[-u_2, -u_1]$. In other words, for the signal f in (4.3) *the reflection about O of $S/\sqrt{(2\pi)}$ gives the distribution of power or intensity over the frequency band*. Naturally, we retain this interpretation of S when f is any arbitrary signal in \mathcal{S} . For signals $f \in \mathcal{S}'$,

$$\{S(\infty) - S(-\infty)\}/\sqrt{(2\pi)}$$

equals $\phi(0)$, the total power of the signal f , cf. (2.5) et seq.; but for signals $f \in \mathcal{S} - \mathcal{S}'$, $\{S(\infty) - S(-\infty)\}/\sqrt{(2\pi)}$ is the power due to *finite* frequencies only, and $\phi(0) - \{S(\infty) - S(-\infty)\}/\sqrt{(2\pi)}$ is that arising from the frequencies $\pm \infty$. Thus *the class $\mathcal{S} - \mathcal{S}'$ comprises precisely those signals which have infinite frequencies in addition to finite frequencies*.

(ii) In the preliminary stages of optics light signals f_1, f_2 are considered to be coherent or interfering when the intensity of the signal $f_1 + f_2$ is “noticeably greater” or “noticeably smaller” than that “ordinarily” observed. The signals are said to be incoherent when the phase difference “varies rapidly and irregularly with time,” so that to the eye or other optical device the result appears as a steady intensity, cf. Rossi (64, p. 119). Wiener explicated this provisional and imprecise concept of coherence by defining f_1, f_2 to be *completely incoherent* when the cross-covariance $\phi_{12}(t) = 0$ for all $t \in (-\infty, \infty)$, and to be *completely coherent* in case $\phi_{12}(t) = \phi_{11}(t)$ for each t . Following him,

¹² In [GHA] Wiener uses the term “energy,” but this is rightly discarded in favor of “power” in later writings, e.g. in [TS, p. 42].

we may interpret ϕ_{12} or a suitable function containing it, such as $\phi_{12}/\sqrt{(\phi_{11}\phi_{22})}$ or the matrices ϕ and S , as *measures of coherence*.

With these definitions Wiener was able to explain why the intensity I of the superposition f_1+f_2 of two completely incoherent signals f_1, f_2 is I_1+I_2 . This fact¹⁸ was puzzling, since Maxwell's equations tell us that the intensity is proportional to $|f(t)|^2$, and obviously $|f_1(t)+f_2(t)|^2 \neq |f_1(t)|^2 + |f_2(t)|^2$. Wiener's explanation went somewhat as follows. Since neither the eye nor any other optical device can resolve the very rapid fluctuations of intensity which occur because of the very large number of independent atomic sources involved, the observed intensity of f is really a time-average $I = (1/2T) \int_{-T}^T |f(t)|^2 dt$. In this average even a short duration T , e.g. 1 microsecond, is so great from the atomic standpoint that the error committed in letting $T \rightarrow \infty$ and replacing I by $\phi(0)$ is insignificant. In effect, the observed intensity of f is $\phi(0)$. Now in case $f=f_1+f_2$ we have

$$\phi(0) = \phi_{11}(0) + \phi_{22}(0) + 2 \text{ real } \phi_{12}(0)$$

which, when the light signals are incoherent, i.e. $\phi_{12}(0)=0$, reduces to $I=I_1+I_2$.

(iii) A light signal with a fixed direction of propagation z at a point r is called *elliptically polarized*, if the tip of $E(t)$ describes an ellipse in the plane $\pi \perp z$. (Circularly and plane polarized lights are obvious special cases.) The signal is said to be *unpolarized* when the tip of $E(t)$ moves "rapidly and erratically" in π . To explicate this interesting but imprecise idea Wiener took an orthonormal basis $B = \{x, y\}$ in π , and represented the light signal relative to B by the 2-ple signal $f_B = (E_x, E_y)$. f_B will of course have a 2×2 matricial covariance function ϕ_B . Wiener defined the light signal to be *completely unpolarized* if for some basis B in π ,

$$(4.5) \quad \phi_B(t) = \phi(t)I, \quad t \in (-\infty, \infty).$$

For any other orthonormal basis B' in π we have obviously

$$\phi_{B'}(t) = U\phi_B(t)U^*$$

where U is an orthogonal matrix. Hence light is completely unpolarized, if and only if the equation (4.5) holds for all orthonormal bases B in π . Obviously the spectral distribution (or coherency) matrix for such a signal will take the form $S(\cdot)I$ in any basis B .

On the other hand, for elliptically polarized monochromatic light of frequency ω , we find that relative to the orthonormal basis formed by the principal axes of the ellipse

¹⁸ Crudely expressed by the statement "two candles are twice as bright as one."

$$\mathbf{f}_B(t) = (a \cos(\omega t + \alpha), b \sin(\omega t + \alpha)),$$

$$\Phi_B(t) = \frac{1}{2} \begin{bmatrix} a^2 \cos \omega t & -ab \sin \omega t \\ ab \sin \omega t & b^2 \cos \omega t \end{bmatrix}.$$

This Φ_B cannot be diagonalised by a (real) rotation of coordinate axes. Wiener proved that any partially polarized monochromatic signal is decomposable into a completely unpolarized signal and an elliptically polarized one. He also discussed the idea of “degree of polarization” and the structure of the group of optical transformations [GHA, §9].

Wiener's ideas have found a place in standard works on optics, e.g. in Chapter X in *Principles of optics* by M. Born and J. Wolf. See also §2 in the paper *Fluctuations of light beams* by L. Mandel in *Progress in Optics*, Vol. II. His ideas appear in such works in a garb which may seem strange to mathematical readers.

5. Signals with absolutely continuous spectra defined by random processes. In [GHA, §§11–13] Wiener gave examples to show that signals in S' can have spectra of all possible types: saltus, singular, or absolutely continuous. Two examples for the last type are derived from stochastic processes, and are especially interesting in pointing to later developments. In our comments we have simplified Wiener's discussion.

EXAMPLE 1 [GHA, §12]. Let \tilde{r}_k be the k th Rademacher function on $[0, 1]$.¹⁴ After Borel we know that the \tilde{r}_k form an independent family of variates on $([0, 1], \mathcal{G}, \text{Leb.})$, where \mathcal{G} is the family of Borel subsets of $[0, 1]$. Consequently the bisequence

$$(r_n)_{n=-\infty}^{\infty} = (\dots, \tilde{r}_6, \tilde{r}_4, \tilde{r}_2, \tilde{r}_1, \tilde{r}_3, \tilde{r}_5, \dots)$$

where $r_n = \tilde{r}_{2(1-n)}$ or \tilde{r}_{2n-1} according as $n \leq 0$ or $n \geq 1$, is an independent, stationary, discrete parameter SP. In essence, Wiener built from this a continuous parameter SP $x(\cdot, \cdot)$ over $([0, 1], \mathcal{G}, \text{Leb.})$, viz.

$$(5.1) \quad x(t, \alpha) = \sum_{k=-\infty}^{\infty} \chi_{J_k}(t) r_k(\alpha), \quad t \in (-\infty, \infty), \quad \alpha \in [0, 1]$$

where $J_k = [k-1/2, k+1/2]$.¹⁵ He showed that for almost all α , the signal or path $x(\cdot, \alpha)$ is in S' , $\phi_\alpha = \phi$ and $S_\alpha = S$, where ϕ is the triangular function

¹⁴ I.e., $\tilde{r}_k(\alpha) = 1 - 2d_k(\alpha)$, $0 \leq \alpha \leq 1$, $k \geq 1$, where $d_k(\alpha)$ is the k th digit in the binary expansion of α .

¹⁵ χ_J denotes the indicator function of J . Actually Wiener took $J_k = [k, k+1)$, but our choice of J_k is a little nicer.

$$\phi(t) = \begin{cases} 1 - |t|, & -1 \leq t \leq 1 \\ 0, & |t| > 1 \end{cases}$$

and S is absolutely continuous on $(-\infty, \infty)$ with

$$(5.2) \quad S'(u) = \sqrt{(2\pi)} |\tilde{\chi}_{J_0}(u)|^2 = \sqrt{\left(\frac{2}{\pi}\right) \frac{1 - \cos u}{u^2}}.$$

To understand this example better, let $\xi(\Delta) = \sum_{k \in \Delta} r_k$ for any Borel subset Δ of $(-\infty, \infty)$. Then ξ is a measure with values in $L_2[0, 1]$ such that $\xi(\Delta_1), \xi(\Delta_2)$ are independent when Δ_1, Δ_2 are disjoint—a so-called *countably additive, independently scattered (c.a.i.s.) measure*. In terms of ξ we can rewrite (5.1) as an equation with stochastic integrals:

$$(5.3) \quad x(t, \cdot) = \int_{-\infty}^{\infty} \chi_{[t-1/2, t+1/2]}(t) \xi(d\tau) = \int_{-\infty}^{\infty} \chi_{J_0}(t - \tau) \xi(d\tau).$$

The process $x(\cdot, \cdot)$ is thus a (2-sided) moving average of the c.a.i.s. measure ξ . But since ξ is concentrated on the set of integers, its variance-measure $|\xi(\cdot)|_2^2$ is not invariant under the translation group on $(-\infty, \infty)$. Hence the process $x(\cdot, \cdot)$ is not stationary. Thus Wiener has exhibited a *nonstationary SP almost all signals (i.e. sample-functions) of which are “stationary”* in the sense of belonging to \mathcal{S}' and having the same covariance and spectral functions ϕ and S .

EXAMPLE 2 [GHA, §13]. The last example shows that the signal resulting from “a haphazard sequence of positive and negative rectangular impulses” almost always has the spectral density S' given in equation (5.2), but that such a signal cannot be derived from a stationary SP. With remarkable insight Wiener saw that to remedy this situation one has “to eliminate the equal spacing of the individual impulses, to reduce the sequence of impulses to such an irregularity as is found in the Brownian motion” (p. 213). To put it in contemporary terms, we must replace the discrete c.a.i.s. measure ξ of Example 1 by the Brownian measure η with values in $L_2[0, 1]$ (built from the increments of the BMSP) for which the spectrum is the entire real line \mathbb{R} and the variance measure $|\eta(\cdot)|_2^2$ is Lebesgue and hence translation invariant.

The replacement of ξ by η led Wiener to the consideration of the SP $y(\cdot, \cdot)$ over $([0, 1], \mathcal{B}, \text{Leb.})$ defined by¹⁶

¹⁶ We are unable to adhere to Wiener's notation. Our W is his ϑ (p. 225), our η is the measure (set-function) induced by his point-function ξ . Our $y(\cdot, \cdot)$ is his f in (13.36). His $\xi(t), f(t)$ should of course be written $\xi(t, \alpha), f(t, \alpha)$, with $t \in (-\infty, \infty)$,

$$(5.4) \quad y(t, \cdot) = \int_{-\infty}^{\infty} W(t - \tau) \eta(d\tau), \quad W \in L_2(-\infty, \infty).$$

By $5\frac{1}{2}$ pages (pp. 226–231) of hard analysis in which a generalization of the Tauberian result (2.7) is used, he was able to show that

$$(5.5) \quad \begin{aligned} & \text{almost every realization } y(\cdot, \alpha) \text{ of the process } y(\cdot, \cdot) \text{ belongs to} \\ & S', \text{ and } S_\alpha \text{ is absolutely continuous on } (-\infty, \infty) \text{ with } S'_\alpha \\ & = \sqrt{(2\pi)} |\tilde{W}(\cdot)|^2. \end{aligned}$$

The significance of this result and a short proof for it became apparent only after Birkhoff proved the ergodic theorem and Paley and Wiener [92, §40] showed that there is a flow $T_t, t \in \mathbb{R}$ such that

$$(5.6) \quad \{\eta(\Delta)\}(T_t \alpha) = [\eta(\Delta + \{t\})](\alpha), \quad \Delta \subseteq \mathfrak{A}, \quad \alpha \in [0, 1],$$

and T_t is measure preserving and ergodic on $([0, 1], \mathfrak{B}, \text{Leb.})$. T_t is referred to as *the flow of Brownian motion or of white noise*. Since it is Lebesgue measure-preserving, the process $y(\cdot, \cdot)$ is a strictly stationary moving average; in fact

$$(5.7) \quad y(t + \tau, \alpha) = y(t, T_\tau \alpha).$$

Hence its covariance function γ and (absolutely continuous) spectral distribution F are given by

$$(5.8) \quad \gamma(t) = \int_{-\infty}^{\infty} W(t - \tau) \overline{W}(-\tau) d\tau, \quad F'(\lambda) = \sqrt{(2\pi)} |\tilde{W}(\lambda)|^2,$$

cf. Doob (20, p. 532 ff). Now consider a realization $y(\cdot, \alpha)$ of the process $y(\cdot, \cdot)$. Since the T_t -flow is ergodic, therefore by Birkhoff's Theorem for almost all α

$$\begin{aligned} \phi_\alpha(t) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y(t + \tau, \alpha) \cdot y(\tau, \alpha) d\tau \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y(t, T_\tau \alpha) \cdot y(0, T_\tau \alpha) d\tau \quad \text{by (5.7)} \\ &= \int_0^1 y(t, \alpha) \cdot y(0, \alpha) d\alpha = \gamma(t). \end{aligned}$$

Hence by (5.8) for almost all α , $S'_\alpha = F' = \sqrt{(2\pi)} |\tilde{W}(\cdot)|^2$.

$\alpha \in [0, 1]$. Wiener uses ξ, f to refer sometimes to the random variable and sometimes to the signal. In place of our (5.4) he has (13.36)

$$y(t, \alpha) = \int_{-\infty}^{\infty} \xi(\tau, \alpha) dW(t + \tau).$$

With his additional assumptions on W this equation reduces essentially to (5.4) on integrating by parts. But the last integral has filter theoretic significance, cf. §23 below.

We see from this that although the flow T_t is not mentioned in [GHA], Wiener's $5\frac{1}{2}$ page proof of (5.5) was in effect a direct proof of the existence (a.e.) of time averages for this flow—a remarkable analytic feat.

II. ASSIMILATION AND CONSOLIDATION (1930–1940)

Wiener did some of his best mathematical work during the thirties. But as far as research in prediction and filter theory is concerned, the period was primarily a preparatory one in which he consolidated his intellectual position. Several strands are discernible in the mathematical activity of this period, which had a profound effect on Wiener's later work, and which therefore call for comment. As before we shall eschew strict chronology.

6. Theory of stochastic processes. First we must mention the epoch-making contribution of Kolmogorov (37) in setting up the theory of stochastic processes on a sound footing. The basic concepts of random variate, conditional probabilities and expectations were cleared up. Wiener used these ideas later in his nonlinear prediction. Kolmogorov's treatment rested on the construction of a probability measure in an infinite dimensional product space, starting from a properly indexed hierarchy of marginally related probability spaces—a “stochastic family” in the sense of Bochner (6). A precursor of this general *Kolmogorov measure* was of course the *Wiener measure* in the space of continuous functions. Wiener was the pioneer, but the Kolmogorov systematization was a boon to all workers in the field.

Next came the formalization and study of the concept of a *stationary* SP due to Khinchine, Cramer, and others. We have already referred to the work of Bochner and Cramer on the Fourier-Stieltjes representability of continuous positive definite functions on $(-\infty, \infty)$ (§2). In 1934 Khinchine (36) showed that a weakly stationary SP has a spectral distribution function, and around 1940 Kolmogorov, Cramer and Loeve (38), (12), (44) showed that such a process admits a “spectral representation.” Discrete-parameter and multivariate extensions of these results were made by H. Wold (66) and Cramer (12). All these contributions were to play an important part in Wiener's later work.

We must next refer to von Neumann's spectral theorem for unitary operators on Hilbert space and Stone's celebrated extension of this to one-parameter unitary groups. As Kolmogorov (38), (40) observed around 1940 these theorems provide an extremely elegant and unified treatment of the spectral theory of weakly stationary SP's. Kolmogorov and Karhunen (35) showed that associated with every *weakly stationary* SP ($f_t, -\infty < t < \infty$) is a unitary group (U_t ,

$-\infty < t < \infty$) such that $f_t = U_t(f_0)$. Here t runs over the set of integers or real numbers according as time is discrete or continuous. Assume that time is continuous, and let E be the spectral measure of the unitary group:

$$U_t = \int_{-\infty}^{\infty} e^{-it\lambda} E(d\lambda).$$

Then¹⁷

$$F(\lambda) = \sqrt{(2\pi)} |E(-\infty, \lambda] f_0|^2, \quad \xi_\lambda = E(-\infty, \lambda) f_0$$

give respectively the spectral distribution F of $(f_t, -\infty < t < \infty)$ and the associated process $(\xi_\lambda, -\infty < \lambda < \infty)$ of orthogonal increments; thus

$$(f_{t+h}, f_t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-ith} dF(\lambda), \quad f_t = \int_{-\infty}^{\infty} e^{-it\lambda} d\xi_\lambda.$$

Similar expressions are available in the discrete case. The results of Khinchine and Cramer thus become easy corollaries of Stone's Theorem, and the study of stationary SP's is reduced to that of *stationary curves or stationary sequences in Hilbert space*.

As Kakutani (32) remarks Wiener did not immediately avail himself of these immense simplifications. [TS] would have been easier on mathematical readers had he done so, though perhaps to engineers it might have been an even greater "yellow peril." However, Wiener adopted the Hilbertian approach in his later papers under the stimulus of his younger collaborators.

7. Ergodic theory. Stone's Theorem was destined to influence Wiener's work in yet another way. It suggested to Koopman the possibility of studying the asymptotic behavior of dynamical systems, governed by measure-preserving flows on a phase space (Ω, \mathcal{B}, P) , by means of the spectral resolution and infinitesimal generator of the induced unitary group on $L_2(\Omega, \mathcal{B}, P)$. This led to the proof of the ergodic theorems of von Neumann, Birkhoff, and others in 1931.

It seems that at first Wiener regarded these theorems merely as useful tools to deal with his time averages. But under the influence of E. Hopf and by the natural evolution of his own thought he soon came to regard the Birkhoff theorem as a mighty beacon which made possible the rigorous construction of statistical mechanics as envisaged by J. W. Gibbs. The fact that ergodicity has to be postulated in order to go from time-averages to the expectations and other well-known averages of probability theory never bothered Wiener, for his

¹⁷ The $\sqrt{(2\pi)}$ in the expression for F is in keeping with the conventions adopted for the spectra of signals in I above. In [185], [186] 2π is used instead of $\sqrt{(2\pi)}$.

scientific philosophy permitted the free creation of bold and ideal hypotheses. In 1938–1939 Wiener gave a unified treatment of different versions of the ergodic theorem and extended them to abelian groups with several generators [117], [108], [128].

8. Hardy class functions. In 1934 Paley and Wiener proved that $f \in L_2(-\infty, \infty)$ is the boundary value of a function f_+ in the Hardy class H_2 on the upper half plane, if and only if its direct Fourier-Plancherel transform \hat{f} vanishes on $(-\infty, 0)$, [92, p. 8]. They also showed that the n.a.s.c. that a function $\phi \geq 0$ a.e. on $(-\infty, \infty)$ and in $L_1(-\infty, \infty)$ be expressible in the form $\phi = |f|^2$ a.e., where f is as just described, is that $\{\log \phi(\lambda)\}/(1+\lambda^2) \in L_1(-\infty, \infty)$, [92, pp. 16–17]. Both results play a central role in the theory of filters. In 1935 came R. Nevanlinna's book (57), in which the earlier work of F. Riesz, Nevanlinna, and Szegö on the canonical factorization of functions in the Hardy classes on the disc appeared as an elegant and coherent theory. Initially Wiener did not feel much need for this powerful theory in his work on prediction. But as he delved deeper into the field, he had to appeal to it.

9. The Hopf-Wiener integral equation. Wiener's interest in integral equations in which the integral is a one-sided convolution was aroused by his colleague E. Hopf. The general H.W. equation of the second kind with unknown f is

$$f(t) + \int_0^\infty K(t-x)f(x)dx = g(t), \quad t \geq 0,$$

the corresponding equation of the first kind (encountered in prediction and filtering) being

$$\int_0^\infty K(t-x)f(x)dx = g(t), \quad t \geq 0.$$

Wiener was struck by this equation. In [177, p. 143] he writes:

The equations for radiation equilibrium in the stars belong to a type now known by Eberhard Hopf's name and mine. They are closely related to other equations which arise when two different physical regimes are joined across a sharp edge or a boundary, as for example in the atomic bomb, which is essentially the model of a star in which the surface of the bomb marks the change between an inner regime and an outer regime; . . .

From my point of view, the most striking use of Hopf-Wiener equations is to be found where the boundary between the two regimes is in time and not in space. One regime represents the state of the world up to a given time and the other regime the state after that time. This is the precisely appropriate tool for certain aspects of the theory of prediction, in which a knowledge of the past is used to determine the future.

As we shall see, this fascination with the H.W. equation had the

interesting effect of leading Wiener to a treatment of prediction, which was at once more limited theoretically and more fruitful practically than that found independently by Kolmogorov (40) in 1941. In Kolmogorov's work the H.W. equation does not occur.

III. NONLINEAR PREDICTION

10. On predictability in semi-exact sciences. In fields such as meteorology the enormity of free coordinates, our ignorance of important factors and the sparseness of our data preclude us from establishing and solving strict dynamical equations. Even so we are able to make predictions. How are we to account for predictability in such "semi-exact" fields? Wiener's position on this important question and on the problem of effecting such prediction is stated in the essay [170]. This appeared in 1954 after much of his specialized work in prediction had been done. But on account of its comprehensive and thought-provoking nature we shall comment on its now, and then turn to specialized problems in prediction.¹⁸

Let Ω be the set of signals pertaining to the field in question, e.g., each $\omega \in \Omega$ may be the record of temperatures at different times (discrete or continuous) at a place P_ω . Assume that time is continuous, so that each ω is a function on $(-\infty, \infty)$, and that $t=0$ represents the present moment. Wiener held that while we cannot in general forecast the future value, $\omega(5)$ say, of a particular signal ω on the basis of our knowledge of the past, we can on this basis forecast the proportion p of the signals ω in Ω for which $\omega(5) \in S$, say, S being a given (Borel) set of numbers. To allow for the possibility that Ω is infinite, p has to be interpreted as a probability. Thus the first premiss of Wiener's theory is that *there exists a probability measure P on a Borel algebra \mathfrak{G} over the signal-space Ω .* Another premiss is that the signals are "not tied down to any specific origin in time" [TS, p. 15], i.e., that Ω is closed under translations T_h : $\{T_h(\omega)\}(t) = \omega(t+h)$, $t, h \in (-\infty, \infty)$.

Now let p_t be the t th coordinate functional on Ω , i.e. $p_t(\omega) = \omega(t)$, and let \mathfrak{G}_t be the Borel subalgebra of \mathfrak{G} generated by the functions p_s , $s \leq t$. Consider a set $A \in \mathfrak{G}_t$, $t > 0$, such that $A \notin \mathfrak{G}_0$; e.g.

$$(10.1) \quad A = \{\omega : \omega(5) \in [a, b] \text{ & } \omega(7) \in [c, d]\} \in \mathfrak{G}_7.$$

Notice that since $T_h^{-1}(\mathfrak{G}_t) = \mathfrak{G}_{t+h}$, we have

¹⁸ [170] is couched in cryptic language. Some of the passages have to be read with considerable empathy. For instance, Wiener rightly emphasizes the central role of ergodicity in his theory (p. 249); but on p. 248, lines 6–8 and p. 250, 3rd paragraph, he gives the misleading impression that prediction can be carried out even when the set of moments accessible to observation is devoid of group structure.

$$T_7(A) = \{\omega : \omega(-2) \in [a, b] \text{ & } \omega(0) \in [c, d]\} \in \mathcal{G}_0.$$

Without further hypotheses $P(A)$ cannot be found from our observations as these extend only up to $t=0$. But assuming that $P(B)$ can be found from observations for $B \in \mathcal{G}_0$ and that the translations T_h preserve P measure, we can find $P\{T_7(A)\}$ from our observations and hence also $P(A) = P\{T_7(A)\}$.

Wiener invoked ergodicity to explain how $P(B)$, $B \in \mathcal{G}_0$, can be found from observations in the past. Assuming that the translations T_h are ergodic, Birkhoff's theorem assures us that for almost all signals ω

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau}^0 \chi_B(T_t \omega) dt = \int_{\Omega} \chi_B(\omega) P(d\omega) = P(B).$$

If $B = T_7(A)$ where A is as in (10.1), then $\int_{-\tau}^0 \chi_B(T_t \omega) dt$ is the Lebesgue measure of the set

$$\{t : -\tau \leq t \leq 0 \text{ & } \omega(t-2) \in [a, b] \text{ & } \omega(t) \in [c, d]\}.$$

This measure can, in principle, be found approximately from our record of the signal values $\omega(t)$ for $t \leq 0$. Thus a third premiss of Wiener's theory is that *the translations T_h , $-\infty < h < \infty$, form an ergodic, measure-preserving flow on (Ω, \mathcal{G}, P)* .

In short, Wiener's prediction theory is based on three postulates: (i) the signal space Ω is closed under all translations T_h , (ii) there exists a probability measure P on a Borel algebra \mathcal{G} over Ω , (iii) the translations T_h , $-\infty < h < \infty$, form a P -measure preserving and ergodic flow.

11. Nonlinear prediction (elementary standpoint). The next problem is to show how prediction is to be carried out on the basis of the postulates (i)–(iii) of §10. Now a probability space (Ω, \mathcal{G}, P) in which Ω consists of signals, i.e. functions on $(-\infty, \infty)$, and satisfies postulate (iii) of §10 is simply a stationary, ergodic SP in its “coordinate representation,” cf. Doob (20, I §6, X §1). As Doob has emphasized such a representation is dispensable. Viewed from a coordinate-free standpoint, *our problem is to carry out the best (nonlinear) prediction of a strictly stationary, ergodic SP*.

Wiener discussed this problem for real SP with discrete time in the joint paper [196] with the writer published in the Harald Cramer Volume. Adopting the RMS error criterion, which is standard in communication theory and in physics, cf. [TS, p. 13], it is shown that for a real-valued strictly stationary SP $(f_n)_{n=0}^{\infty}$ over (Ω, \mathcal{G}, P) such that $E(f_n) = 0$ the best prediction of f_n with lead ν is the conditional expectation $E(f_n | \mathcal{G}_0)$, where \mathcal{G}_0 is the Borel subalgebra of \mathcal{G} spanned

by the f_k for $k \leq 0$. Assuming that f_0 and therefore each f_k is in $L_\infty(\Omega, \mathcal{B}, P)$, we show that $E(f_v | \mathcal{G}_0)$ is the orthogonal projection of f_v on the L_2 -closure of the linear algebra \mathcal{Q}_0 spanned by the f_k , $k \leq 0$, the so-called "nonlinear past of f_0 ." Finally, with the additional assumption that for distinct integers n_1, \dots, n_q the spectra of the distribution functions of the q -variates $(f_{n_1}, \dots, f_{n_q})$ have positive q -dimensional Lebesgue measure, it is proved that

$$E(f_v | \mathcal{G}_0) = \lim_{n \rightarrow \infty} Q_n \{f_0, f_{-1}, \dots, f_{-m_n}\}$$

where m_n is a non-negative integer depending on n , and the Q_n are real polynomials in $m_n + 1$ variables, the coefficients of which are computable expressions of the moments of the SP. These moments can in principle be determined from time series data in the past on the basis of ergodicity.

Thus, under rather natural restrictions the nonlinear prediction problem is solvable, and indeed reducible to a linear problem, viz. the determination of the orthogonal projection of f_v on the "nonlinear past," i.e. on a well-defined subspace of the Hilbert space $L_2(\Omega, \mathcal{B}, P)$. Of course, the practical difficulties of carrying out this solution are enormous, cf. §12.

12. Nonlinear prediction (advanced standpoint). The solution just outlined has the shortcoming of requiring the inversion of larger and larger matrices to get the different Q_n . Wiener felt that a better approach to nonlinear prediction required a deeper analysis of strictly stationary processes. In the linear case the corresponding analysis is that due to Wold and Kolmogorov. The problem now was to carry out a similar, but nonlinear, time-domain analysis of strictly stationary SP, and to follow this up with a characteristic functional analysis.

Wiener seems to have begun efforts in this direction around 1953, being guided to some extent by his earlier work on homogeneous chaos [108]. In 1956 he worked on the problem with G. Kallianpur at Calcutta. Their results appear in a technical report¹⁹ and also in extremely diffuse form, and unfortunately without reference to Kallianpur, in the chapters entitled Coding and Decoding in Wiener's book [NPRT] published in 1958.²⁰ We must now examine this compelling theory, although it contains a lacuna as we shall see.

¹⁹ Nonlinear prediction (with G. Kallianpur), Technical Report No. 1 (1956), Office of Naval Research, Cu-2-56-Nonr-266, (39)-CIRMP, Project NR-047-015.

²⁰ The material also appears in a mimeographed write-up by E. J. Akutowicz of lectures given by Wiener at Massachusetts Institute of Technology after his return from India. This version differs from the Kallianpur-Wiener report (see footnote 19) in proof techniques but not in essential conception and results.

Wiener and Kallianpur consider a real-valued strictly stationary SP $(f_n)_{-\infty}^{\infty}$ over (Ω, \mathcal{G}, P) . For any set J of integers let \mathcal{G}_J be the Borel algebra spanned by f_j , $j \in J$ and \mathcal{F}_J be the family of Borel sets in the (not necessarily finite-dimensional) Euclidean space \mathcal{R}^J , \mathcal{R} being the real number field.²¹ Introduce the abbreviations

$$\mathcal{G}_n = \mathcal{G}_{\{n\}}, \quad \mathcal{F}_n = \mathcal{F}_{\{n\}}, \quad \text{where } J = \{n, n-1, \dots\}.$$

The (microscopic) conditional probability $P(f_0^{-1}(-\infty, \lambda] | \mathcal{G}_{-1})$ is a \mathcal{G}_{-1} -measurable function on Ω . Hence (cf. e.g. Dynkin (21, 1.5)) there exists a \mathcal{F}_0 -measurable function G on $\mathcal{R}^{\{0,-1,-2,\dots\}}$ such that for each real λ , $G(\lambda, \cdot)$ is \mathcal{F}_{-1} -measurable and

$$P(f_0^{-1}(-\infty, \lambda] | \mathcal{G}_{-1})(\omega) = G(\lambda, f_{-1}(\omega), f_{-2}(\omega), \dots).$$

Wiener's first hypothesis is that for almost all $\omega \in \Omega$,

(A) $G(\cdot, f_{-1}(\omega), f_{-2}(\omega), \dots)$ is strictly increasing and continuous on \mathcal{R} .

This condition entails, of course, that $\mathcal{G}_{-1} \neq \mathcal{G}_0$, i.e. that the SP $(f_n)_{-\infty}^{\infty}$ is nondeterministic. Let

$$(12.1) \quad g_0(\omega) = G(f_0(\omega), f_{-1}(\omega), f_{-2}(\omega), \dots).$$

Obviously the range of g_0 is essentially contained in $[0, 1]$. It can be shown that g_0 is independent of f_{-1}, f_{-2}, \dots and is uniformly distributed. Letting Φ be the $(0, 1)$ normal distribution function on \mathcal{R} and $h_0 = \Phi^{-1}(g_0)$ it follows that h_0 is $(0, 1)$ normally distributed and independent of f_{-1}, f_{-2}, \dots , and

$$h_0(\omega) = H(f_0(\omega), f_{-1}(\omega), \dots), \quad H = \Phi^{-1} \circ G.$$

Defining

$$(12.2) \quad h_n(\omega) = H(f_n(\omega), f_{n-1}(\omega), \dots), \quad -\infty < n < \infty,$$

it follows that $(h_n)_{-\infty}^{\infty}$ is an independent, $(0, 1)$ normally distributed SP, and h_n is \mathcal{G}_n -measurable and independent of f_{n-1}, f_{n-2}, \dots . We can also show that $\mathcal{G}_n = \mathcal{G}(h_n, f_k, k < n)$. Since

- (i) h_n is $(0, 1)$ normally distributed,
- (ii) h_n is \mathcal{G}_n -measurable and independent of f_{n-1}, f_{n-2}, \dots ,
- (iii) $\mathcal{G}_n = \mathcal{G}(h_n, f_{n-1}, f_{n-2}, \dots)$,

we are entitled to regard h_n as the " n th innovation" of the SP $(f_n)_{-\infty}^{\infty}$. Wiener and Kallianpur thus established the existence of (*non-*

²¹ \mathcal{F}_J is the Borel algebra spanned by all sets of the form $X_{i \in J} B_i$, where B_i are Borel subsets of \mathcal{R} .

(linear) innovations for a stationary SP obeying the condition (A) of strong nondeterminism.²²

Wiener's next objective was to seek hypotheses which would ensure that his original SP $(f_n)_{-\infty}^{\infty}$ be a nonlinear, one-sided moving-average of his innovation SP:

$$(12.4) \quad f_n = F(h_n, h_{n-1}, \dots), \quad -\infty < n < \infty.$$

An easy generalization of 12.3(iii) is

$$(12.5) \quad \mathfrak{G}_n = \mathfrak{G}(h_n, h_{n-1}, \dots, h_{n-k-1}, f_{n-k-2}, \dots), \quad n, k \geq 1.$$

Wiener and Kallianpur now impose their second hypothesis, viz. the SP $(f_n)_{-\infty}^{\infty}$ is *purely nondeterministic*, i.e.

$$(B) \quad \mathfrak{G}_{-\infty} = \bigcap_{n \geq 0} \mathfrak{G}_{-n} = \{\text{void set}, \Omega\}.$$

From (12.5) and (B) they conclude (letting $k \rightarrow \infty$) that $\mathfrak{G}_n = \mathfrak{G}(h_k, k \leq n)$, and thence (12.4).

In 1959, about a year after the publication of [NPRT], M. Rosenblatt (63) gave an example in which the hypotheses (A) and (B) are fulfilled but (12.4) fails. The inference of $\mathfrak{G}_n = \mathfrak{G}(h_k, k \leq n)$ from (12.5) and (B) is thus incorrect. But once (12.4) is properly established (say by suitably strengthening (A) and (B), or by changing the definition of g_0) we would have a theory useful for prediction. Thus with $f_k \in L_2(\Omega, \mathfrak{G}, P)$ and the RMS criterion we would get

$$E(f_n | \mathfrak{G}_0) = H(0, \underset{n \text{ times}}{\dots}, h_0, h_{-1}, \dots), \quad n \geq 0.$$

Once H is known our predictor would be determined. Indeed we could get from (12.4) any statistical parameter of the conditional distribution of f_n relative to \mathfrak{G}_0 , e.g. the median.

13. Unsolved questions. Wiener has bequeathed to posterity the important problem of strengthening his hypotheses (A), (B), (§12) or of changing the definition (12.1) of g_0 so as to ensure both (12.2) and (12.4). Also left to us is the extension of this theory to the continuous parameter case. Here the absence of an atomic time-unit makes the problem of defining nonlinear innovations extremely hard; obviously all we may expect are virtual or differential innovations.

Also remaining on the agenda is the implementation of Wiener's idea expressed on the last page of his brilliant essay [170] of affecting

²² Wiener and Kallianpur seem to have been unaware that their technique for defining g_0 is an interesting "infinite" adaptation of one used by P. Levi (43, Ch. VI) in deriving from a sequence $(f_n)_{1}^{\infty}$ a sequence of independent uniformly distributed varieties g_n such that $f_n = F_n(g_1, \dots, g_n)$.

a nonlinear prediction of a (simple) SP by carrying out a linear prediction of a suitably derived infinite dimensional SP. There also remains the job of supplementing nonlinear, time-domain analysis by a characteristic functional analysis.

IV. LINEAR PREDICTION THEORY

14. The linear prediction and filtering problems. When the strictly stationary SP ($f_k, -\infty < k < \infty$), where k can be integral or real, is Gaussian with zero expectations and $f_k \in L_2(\Omega, \mathcal{G}, P) = \mathcal{H}$, the best (nonlinear) predictor $E(f_{\nu} | \mathcal{G}_0)$, $\nu > 0$, considered in §11, turns out to be the orthogonal projection \hat{f}_{ν} of f_{ν} on the (closed, linear) subspace \mathcal{M}_0 of \mathcal{H} spanned by f_k , $k \leq 0$. When $(f_k)_{-\infty}^{\infty}$ is not Gaussian but merely stationary, \hat{f}_{ν} provides the first, i.e. linear, approximation to the best prediction of f_{ν} . On both counts the problem of finding the orthogonal projection of f_{ν} on \mathcal{M}_0 is important. This is the *linear prediction problem* with lead ν , independently conceived and studied by Wiener and Kolmogorov.

When the time-signals of our stationary SP ($f_k, -\infty < k < \infty$) are contaminated messages we may at the linear level assume that $f_k = f_k^1 + f_k^2$, where the variates f_k^1, f_k^2 represent the (*pure*) *message* and (*pure*) *noise*, respectively. Our problem is to find the best linear approximation to the message-variate f_{ν}^1 in terms of the signal variates f_k , $k \leq 0$, i.e. the orthogonal projection \hat{f}_{ν}^1 of f_{ν}^1 on the subspace \mathcal{M}_0 spanned by f_k , $k \leq 0$. This is the *linear filtering problem* (with lead or lag ν), also originally conceived by Wiener.

How Wiener was drawn to these problems by his involvement with antiaircraft fire control and noise filtration in radar during the war, and how his mathematical background fitted him ideally for this task are recounted in his [204], [177], [TS]. His work in this field falls rather naturally into two periods. The first, 1940–1943, culminated in his [TS], which was completed in early 1942 but appeared in declassified form only in 1949. The second period, 1949–1959, began with his addresses [143], [156] to the CNRS in Paris, and the International Congress of Mathematicians at Harvard, and ended with his papers on multivariate prediction.

During the first period Wiener was interested chiefly in getting autoregressive integral or series representations for the predictor and filter. He used variational techniques and solved the resulting Hopf-Wiener type integral equations. He accomplished all this without the use of too much theory by operating at what Yaglom [71, p. vii] has aptly called “a heuristic level of rigor.” Some of the problems he tackled were formidable, but his work in this period lacked the theoretical strength and completeness of that of Kolmogorov (40).

On the other hand during the second period Wiener adopted a more theoretical approach. He undertook time-domain and spectral analysis, leaning more and more on abstract theory, especially Hilbert spaces.

15. Linear prediction and filtering (first period). Despite its hard mathematical content, which earned it the nickname "the yellow peril," [TS] has had a wide influence in engineering circles. This stemmed from its wealth of engineering insights and its emphasis on the quick solution of engineering problems without much fuss over rigor. The urgencies of war as well as Wiener's long-standing fascination with such problems and the functional equations to which they led had much to do with this attitude. As there is still some doubt as to the locus standi of the prediction and filtering techniques used in [TS] we shall briefly review them here in the light of the current theory of stochastic processes.

Consider a stationary, purely nondeterministic SP (f_t , t real), $f_t \in L_2(\Omega, \mathcal{B}, P)$. As indicated in §10, we can estimate the covariances $\phi(t) = (f_t, f_0)$ from time-signal data in the past on the hypothesis of ergodicity, which Wiener freely made. Thus the covariance function ϕ is "known." Our problem is to find for a given $h > 0$, the orthogonal projection \hat{f}_h of f_h on \mathfrak{M}_0 , the closed subspace spanned by f_t , $t \leq 0$. Wiener studied this problem in the mathematically restrictive but practically important case in which \hat{f}_h is given by an autoregressive Stieltjes integral:

$$(15.1) \quad f_h = \int_0^\infty f_{-r} dw(r)$$

where w is of bounded variation on $[0, \infty)$. (As yet, no nice necessary and sufficient criterion for the validity of such a representation of \hat{f}_h is known.) Since $f_h - \hat{f}_h \perp f_{-t}$, $t \geq 0$, we readily get on taking inner products

$$(15.2) \quad \phi(t+h) = \int_0^\infty \phi(t-\tau) dw(\tau), \quad t \geq 0.$$

This is a *Hopf-Wiener Stieltjes integral equation of the first kind*, by solving which the "unknown" function w is to be found. Wiener arrived at this equation [TS, (2.021)] by a longer variational approach.

To solve (15.2) let us proceed heuristically. We extend w to $(-\infty, \infty)$ by defining it to be zero on $(-\infty, 0)$ and define

$$(15.3) \quad \begin{aligned} \Psi_h(t) &= \phi(t+h) \chi_{[0,\infty)}(t), \quad -\infty < t < \infty, \\ u(t) &= \int_{-\infty}^\infty \phi(t-\tau) dw(\tau) - \Psi_h(t), \quad -\infty < t < \infty. \end{aligned}$$

Denoting as usual the indirect Fourier transform by a tilde, noting that $\tilde{\phi}(\lambda) = F'(\lambda)$, where F' is the spectral density of the SP, and letting

$$(15.3') \quad W(\lambda) = \int_0^\infty e^{i\lambda t} dw(t),$$

we get from (15.3)

$$(15.4) \quad \tilde{u}(\lambda) = F'(\lambda)W(\lambda) - \tilde{\Psi}_h(\lambda), \quad -\infty < \lambda < \infty.$$

Now Wiener knew that his hypothesis that the SP is *nondeterministic*, i.e. $f_h \notin \mathfrak{M}_0$, entails the condition

$$(15.5) \quad \{\log F'(\lambda)\}/(1 + \lambda^2) \in L_1(-\infty, \infty).$$

(Actually the equivalence of (15.5) to nondeterminism was proved only later by Karhunen (35).) From (15.5) and the Paley-Wiener Theorem [92, pp. 16–17]

$$(15.6) \quad F'(\lambda) = |\Phi(\lambda)|^2 \text{ a.e. on } (-\infty, \infty),$$

where Φ has a holomorphic extension to the upper half plane Δ_+ , this extension being in the Hardy class H_2 . One can choose a Φ which is free from zeros in Δ_+ . Then $1/\Phi$ will be holomorphic on Δ_+ and $1/\bar{\Phi}$ on the lower half plane Δ_- . Also since $u=0$ on $[0, \infty)$ and $w=0$ on $(-\infty, 0]$, we see that \tilde{u} and W have holomorphic extensions to Δ_- , Δ_+ respectively. Hence on dividing by $\bar{\Phi}$ in (15.4) we find that

$$\Phi(\lambda)W(\lambda) = [\tilde{\Psi}_h(\lambda)/\bar{\Phi}(\lambda)]_+$$

where $[\]_+$ denotes the operation of cutting off the negative frequencies.²³ It easily follows that

$$(15.7) \quad W(\lambda) = \frac{1}{\Phi(\lambda)} [\tilde{\Psi}_h(\lambda)/\bar{\Phi}(\lambda)]_+ = \frac{1}{\Phi(\lambda)} [e^{-i\lambda}\Phi(\lambda)]_+.$$

This yields W , from which w or a suitably normalized version of w can be retrieved by inversion:

$$w(t) - w(0) = \lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A \frac{1 - e^{-it\lambda}}{i\lambda} W(\lambda) d\lambda.$$

The exact hypotheses needed to validate this heuristic solution have not been discovered. Wiener showed that it goes through in case the spectral density F' is a rational function P/Q with $\deg P \leq \deg Q$. This case, though mathematically restrictive, is important in applications.

²³ I.e. $[\int_{-\infty}^{\infty} e^{it\lambda} g(t) dt]_+ = \int_0^{\infty} e^{it\lambda} g(t) dt$.

The above-mentioned treatment of the prediction problem was in marked contrast to that of Kolmogorov which covered the entire field of prediction of discrete parameter univariate processes, first in the time-domain with heavy emphasis on the fundamental one-sided moving average representation discovered by Wold (66) in 1938, and then in the spectral domain by exploiting powerful theorems on Hardy class functions. Kolmogorov showed that there is an isomorphism between the time and spectral domains. It may be shown that if we define W not by (15.3') but by (15.7), taking the Φ therein to be the *optimal* (or outer) function satisfying (15.6), then \hat{f}_h and W are isomorphs, cf. e.g. [186, 4.11]. Yaglom, Darlington and other workers interested in the engineering side have developed algorithms for computing the predictor in the frequency domain starting directly from this fact and without assuming the autoregressive representation (15.1) in the time-domain. The Hopf-Wiener equation (15.2) is thus avoided (68), (16).

Wiener approached linear filtering in much the same spirit as linear prediction. Let $f_t = f_t^1 + f_t^2$, where f_t^1, f_t^2 represent the pure message and pure noise, respectively. As before we assume as known the auto- and cross-covariance functions ϕ, ϕ_1 :

$$\phi(t) = (f_t, f_0), \quad \phi_1(t) = (f_t^1, f_0).$$

As in (15.1) Wiener assumed that for any real h , the projection \hat{f}_h^1 of f_h^1 on the subspace \mathfrak{M}_0 spanned by $f_t, t \leq 0$ is given autoregressively:

$$(15.1') \quad \hat{f}_h^1 = \int_0^\infty f_{-\tau} dw(\tau).$$

Instead of (15.2) he now gets the Hopf-Wiener type equation

$$(15.2') \quad \phi_1(t+h) = \int_0^\infty \phi(t-\tau) dw(\tau), \quad t \geq 0,$$

cf. [TS, (3.20), (3.52)]. Keeping to our former notation except for using ϕ_1 instead of ϕ in defining $\Psi_h(t)$, we get as before

$$(15.7') \quad W(\lambda) = \frac{1}{\Phi(\lambda)} [\tilde{\Psi}_h(\lambda)/\bar{\Phi}(\lambda)]_+ = \frac{1}{\Phi(\lambda)} [e^{-ih\lambda} F_1(\lambda)/\bar{\Phi}(\lambda)]_+,$$

where F_1 is the cross-spectral distribution of the f_t^1 - and f_t -processes. From (15.7') w can be retrieved under suitable assumptions. Wiener's approach was again rather pragmatic. As in prediction one may show without recourse to (15.1') and (15.2') that the function W in (15.7') is the spectral isomorph of the filtration \hat{f}_h^1 .

16. Linear prediction and filtering (second period). As mentioned in §14, around 1949 Wiener began to veer towards a more theoretical approach to weakly stationary SP's. In his addresses [143, 156] he dealt with moving averages in the time-domain (unfortunately without reference to Wold) and followed this by spectral analysis. This trend persists in the papers [172, 176] in which Hilbert space techniques are increasingly used. But Wiener's analysis was incomplete. For instance, the remote past and nonabsolutely continuous spectra were not considered. Indeed, so thorough had been Kolmogorov's treatment of univariate prediction in the discrete case (40) that there was little left to do.

In [156, 176] Wiener also discussed the prediction of continuous parameter, weakly stationary processes. This aspect of his work is best broached by a quotation [176, p. 183]: "Let the unitary transformation T^{-1} carry $f(\alpha)$ into $g(\alpha)$ if and only if for all n

$$\int_{-\infty}^{\infty} f(S^{-t}\alpha) l_n(t) dt = \int_{-\infty}^{\infty} g(S^{-t}\alpha) l_{n-1}(t) dt.$$

In this interesting but rather cryptic sentence the l_n are Laguerre functions and $(S^t, t \text{ real})$ is a measure-preserving flow on $([0, 1], \text{Borel, Leb.})$. When deciphered the sentence reads: "Let T be the Cayley transform of H , where iH is the infinitesimal generator of the unitary group $(S_t, t \text{ real})$ induced by the S^t -flow: $S_t(f) = f \circ S^t$." It seems that Wiener did not realise this; even so he took the important step of associating with his continuous parameter SP $(S_t(f), t \text{ real})$ the discrete parameter SP $(T^n(f), n = \text{integer})$, and so bringing the discrete theory to bear on the continuous. But he did not pursue this fine idea systematically, and his work on the continuous case during this period is on the whole rather sketchy.

In [143, 176] Wiener also considered linear filtering with discrete time. He employed moving averages and spectral theory to prove results obtained either heuristically or under restrictions in earlier work.

17. Multivariate prediction. No sooner had Wiener tackled the prediction and filtering problems for univariate processes, he turned to the corresponding problems for q -variate processes $(\mathbf{f}_t, t \text{ real})$, where each \mathbf{f}_t is a q -ple vector $(f_i^t)_{i=1}^q$ with $f_i^t \in L_2(\Omega, \mathcal{G}, P)$, and the Gram matrix $(\mathbf{f}_s, \mathbf{f}_t) = [(f_s^i, f_t^i)] = \phi(s-t)$ depends only on the difference $s-t$. In [TS, Ch. IV] he assumed, in keeping with his approach for the case $q=1$, that the predictor $\hat{\mathbf{f}}_h$ has an autoregressive Stieltjes integral representation. He then got instead of the single Hopf-Wiener equation (15.2) a system of q such equations in q unknowns

[TS, (4.105)]. This system yields on solution the first component of \hat{f}_h . To get \hat{f}_h itself we must solve a system of q^2 such equations in q^2 unknowns, i.e. a single matrix Hopf-Wiener equation

$$\phi(t + h) = \int_{-\infty}^{\infty} \phi(t - \tau) d\mathbf{w}(\tau)$$

where \mathbf{w} is the (unknown) $q \times q$ matricial weighting. Wiener's treatment of the solution of this equation was again largely heuristic. It can be carried through only under rather severe restrictions on the process, which, however, are fulfilled in many cases of practical interest.

A more theoretical approach to multivariate prediction occurs in Wiener's address [156], and is amplified in the paper [172] dedicated to Plancherel. Wiener considered a discrete parameter, purely non-deterministic, bivariate SP of full rank, i.e. one for which $q=2$, the so-called remote past $M_{-\infty}$ is $\{0\}$ and the 1-step prediction error matrix has rank 2. By alternating projections in Hilbert space Wiener obtained expressions for the innovation vectors. But questions of existence and those of computation were not kept apart, and this delimited the work. Thus Wiener was able to show that such a process has a spectral density F' and that F' is factorizable:

$$(17.1) \quad F' = \Phi \Phi^*, \quad \text{a.e.,} \quad \Phi \sim \sum_{k=0}^{\infty} C_k e^{ki\theta} \in L_2(C),$$

where C is the unit circle $\{z: |z|=1\}$. But he was not quite able to show that the condition

$$(17.2) \quad \log \det F' \in L_1(C)$$

suffices for such factorization.²⁴ In [172] Wiener also discussed a similar factorization for unitary matrix-valued functions, but his proof is incorrect and the result itself is in doubt.

The work done up to this point by Wiener and others had not cleared up the basic questions of q -variate prediction theory, and clearly pointed to the need for a systematic study of these questions unencumbered by algorithmic considerations. Wiener began this work in collaboration with the writer in 1955–1956 at Calcutta. This research [185, 186], deals with the general case first and only later with special cases.²⁵ As the subject is rather technical we shall assume in what follows that the reader is familiar with the basic concepts.

²⁴ His argument fails when the angle between the past subspaces of the component processes is zero.

²⁵ [185] owes much to Zasuhin's brilliant Doklady note (72), which announced a chain of general results on q -variate processes.

[185] begins with the time-domain analysis of q -variate processes of rank ρ , $1 \leq \rho \leq q$. Following Zasuhin ρ is defined to be the rank of the 1-step prediction error Gram matrix

$$\mathbf{G} = (\mathbf{g}_0, \mathbf{g}_0) = [(g_0^i, g_0^j)]$$

where $\mathbf{g}_0 = (g_0^i)_{i=1}^q$ is the 0th innovation vector. The Wold or rather *Wold-Zasuhin decomposition* is established, and necessary and sufficient criteria given for pure nondeterminism, i.e. for $\mathbf{M}_{-\infty} = \{\mathbf{0}\}$. The paper also contains spectral analysis. Perhaps the most fundamental result is the *determinantal extension of the Szegö-Kolmogorov identity*:

$$(17.3) \quad \log \det \mathbf{G} = \frac{1}{2\pi} \int_0^{2\pi} \log \det \mathbf{F}'(e^{i\theta}) d\theta,$$

where \mathbf{F} is the $q \times q$ matricial spectral distribution of the SP. This result, first stated by Whittle (65), is proved by exploiting theorems on the Hardy classes as well as the concavity of the functional $\log \det$ on the space of $q \times q$ non-negative, hermitian matrices. (We get a generalized Jensen inequality.) From (17.3) it follows at once that (17.2) is the n.a.s.c. that $\rho = q$, i.e. that the SP be of *full rank*. We show next that when $\rho = q$ the absolutely continuous and nonabsolutely continuous parts $\mathbf{F}_a, \mathbf{F}_b$ of \mathbf{F} are the spectral distributions of the purely nondeterministic part $(u_n)_{-\infty}^{\infty}$ and the deterministic part $(v_n)_{-\infty}^{\infty}$ of the Wold-Zasuhin decomposition of $(f_n)_{-\infty}^{\infty}$ —the so-called *concordance of Wold-Zasuhin and Lebesgue-Cramer decompositions*. By appeal to Cramer's criterion that a matrix-valued function be the spectral distribution of a SP (12), we also show that if $\mathbf{F}' \geq 0$, $\mathbf{F}' \in L_1(C)$ and satisfies (17.2), then \mathbf{F}' admits a factorization of the form (17.1), where moreover the factor Φ is *optimal* (“outer” in Beurling's terminology), i.e.,

$$(17.4) \quad \mathbf{C}_0 \geq 0, \quad \log \det \mathbf{C}_0 = \frac{1}{2\pi} \int_0^{2\pi} \log \det \mathbf{F}'(e^{i\theta}) d\theta.$$
²⁶

In [186] an isomorphism is established between the time and spectral domains of a q -variate purely nondeterministic SP of full rank q . But the primary goal is the derivation of an algorithm for the linear prediction $\hat{\mathbf{f}}_\nu$ of \mathbf{f} , with lag ν . $\hat{\mathbf{f}}_\nu$ turns out to be the isomorph of the function \mathbf{Y}_ν :

$$(17.5) \quad \mathbf{Y}_\nu(e^{i\theta}) = [e^{-\nu i\theta} \Phi(e^{i\theta})]_0 + \Phi^{-1}(e^{i\theta})$$

²⁶ Some of these results were obtained independently by M. Rosenblatt, Helson and Lowdenslager, and Iu. A. Rosanov.

where Φ is the unique function satisfying (17.1) and (17.4). The problem is therefore to find an algorithm for the computation of this “optimal factor” of F' . In case $q=1$

$$\Phi(z) = \exp \left\{ \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} - z}{e^{i\theta} + z} \log F'(e^{i\theta}) d\theta \right\}, \quad |z| < 1.$$

But for $q > 1$ no such closed-form expression involving the log is available since matrix multiplication is noncommutative, and all one can hope for is some iterative procedure for finding Φ . Our algorithm emerges when Wiener's approach for $q=2$ based on alternating projections [172] is judiciously fused with a technique of noncommutative factorization used earlier by the writer (46). Under the assumptions $\lambda I \leq F'(e^{i\theta}) \leq \lambda' I$, $0 < \lambda \leq \lambda' < \infty$, we show that

$$(17.6) \quad \sqrt{G}\Phi^{-1} = I - M_+ + (M_+M)_+ - \{(M_+M)_+M\}_+ + \dots,$$

where

$$M = \frac{2}{\lambda + \lambda'} F' - I$$

and the subscript $+$ denotes the operation of cutting off all but positive frequencies. For \hat{f} , itself we obtain a mean-convergent auto-regressive series

$$(17.7) \quad \hat{f}_r = \sum_{k=0}^{\infty} E_{rk} f_{-k},$$

where the E_{rk} are finite sums of the Fourier coefficients of Φ^{-1} and Φ .

A natural sequel to the foregoing study is that of purely nondeterministic processes of degenerate rank: $1 \leq \rho < q$. In this case F is absolutely continuous,

$$(17.8) \quad \det F' = 0, \text{ a.e.}, \quad F' = \Phi\Phi^*, \text{ a.e.}, \quad \Phi \sim \sum_{k=0}^{\infty} C_k e^{ki\theta} \in L_2(C).$$

What conditions must F' satisfy in order that such a “*degenerate rank*” factorization be possible? In [197] Wiener and the writer gave a complete answer for the bivariate case, $q=2$: with $F = [F_{ij}]$,

$$(17.9) \quad \begin{aligned} \log F'_{ii} &\in L_1(C), \quad i = 1 \text{ or } 2, \\ F'_{ji}/F'_{ii} &= \text{radial limit of a beschränktartige function, } j \neq i. \end{aligned}$$

At about the same time Wiener wrote a paper [194] with Akutowicz in which the “full rank” factorization theorem of [185] is reproved *ab initio*. The difficulty which Wiener had encountered in his earlier

work [172, 176] is avoided by abandoning the method of alternating projections in favor of a technique suggested by the time-domain analysis of q -variate processes given in [185].

18. The continuation of Wiener's work on prediction. The work of Kolmogorov and Wiener on prediction has had widespread influence. Activity in the field has been especially vigorous from the time when Wiener's work on multivariate prediction and that of Rosanov (60) in Russian appeared. In the following brief account of these developments we shall only mention work closely related to Wiener's. Our order will be logical rather than chronological.

During 1947–1950 Karhunen (35) and Hanner (26) extended Kolmogorov's work to univariate continuous parameter processes, and thereby settled important questions which Wiener had bypassed in his early, heuristic attacks.²⁷ But the techniques used were somewhat ad hoc in nature. Now (cf. §16) Wiener [176] had suggested the possibility of associating with the weakly stationary, continuous parameter SP (U, f, t real) the weakly stationary, discrete parameter SP ($V^n f, n = \text{integer}$), where V is the Cayley transform of H , iH being the infinitesimal generator of the unitary group (U, t real). J. Robertson and the writer (55) showed that such association results in a coherent and simple development of the entire theory. In particular, the troublesome process of orthogonal increments (ξ_t, t real), the differentials of which are the (virtual) innovations of the f_t -process, is easily obtained.

On the engineering side, Bode and Shannon (7) simplified Wiener's early version of prediction and filtering [TS] using circuit theory concepts. The circuit theory point of view also led Darlington (16) to a simpler form of the theory adaptable to applications exemplifying rational spectra. Zadeh and Ragazinni (72) modified the Wiener theory to cover the case in which the known data is confined to a bounded time-interval in the past. Yaglom (68) put some of this work on a rigorous footing. J. Chover (9), (10) made an interesting mathematical analysis of the case in which such prediction in terms of a bounded interval is representable autoregressively by a Stieltjes integral, cf. (15.1). Dolph and Woodbury (18) studied the predictor and filter, again on the basis of a bounded time-interval, for SP's, the signals of which are generated by driving linear differential systems by white noise. Kalman and Bucy (34) also considered this problem, but from the general standpoint of stochastic control theory.

To turn to the q -variate theory, the isomorphism between the time

²⁷ The state of the subject at this point (1950) is well exposed in Doob's book (20, Ch. XII).

and spectral domains established in [186] under the hypothesis of full rank and pure nondeterminism was obtained without restriction by M. Rosenberg (59) and Rosanov (62'). Research in the field was thereby freed from reliance on unnecessary hypotheses. Matveev (56) studied degenerate rank factorizations of matrix-valued functions for any $q \geq 1$. His condition for the factorization (17.8) with rank $\Phi = \rho$, $1 \leq \rho < q$ is a direct generalization of the condition (17.9) proved in [197] for $q = 2$, $\rho = 1$. Somewhat less tractable conditions for degenerate rank factorization were found by Helson and Lowdenslager (28). The writer (47) showed that the concordance of the Wold-Zasuhin and Lebesgue-Cramer decompositions, established in [185] in the full rank case, breaks down when $1 \leq \rho < q$. For the case $q = 2$, $\rho = 1$, he gave necessary and sufficient conditions for the prevalence of concordance, which J. Robertson (58) generalized to $q \geq 1$, $1 \leq \rho < q$, by first proving an elegant result on the ranges of the matrices $F'_x(e^{i\theta})$, $F'_y(e^{i\theta})$, $F'_z(e^{i\theta})$, where the x_n -process is the sum of orthogonally related processes y_n and z_n . Some of Robertson's results were duplicated independently in China by Jang Ze-pei (31).

On the computational side, the writer (48) showed that the algorithm (17.6) for finding the optimal factor Φ of F' satisfying (17.1) and (17.4) extends to the case in which the reciprocal matrix $(F')^{-1}$ and the quotient $\lambda'(\cdot)/\lambda(\cdot)$ of the largest to the smallest eigenvalue of F' are in $L_1(C)$. He also showed that the autoregressive series (17.7) for the predictor is available under the weak hypotheses $F' \in L_\infty(C)$, $(F')^{-1} \in L_1(C)$. Yaglom (70) developed algorithms for prediction for continuous parameter q -variate processes with rational spectra.

Gangoli (24) developed a q -variate prediction theory valid for $q = \infty$. He defined a SP as a bisequence of bounded linear operators F_n from Hilbert spaces \mathcal{H} to \mathcal{H}' , where $\dim \mathcal{H} = q$, and showed that many of the definitions, result and proofs given in [185] carry over to the case $q = \infty$. Gangoli also considered the factorization $W' = \Phi \Phi^*$, where W' is a function on the unit circle C , the values of which are non-negative, hermitian operators on \mathcal{H} to \mathcal{H} , and Φ has a one-sided Fourier development. This factorization had been treated earlier by Devinatz (17), who showed that a sufficient condition is that $\log \lambda(\cdot) \in L_1(C)$, where $\lambda(e^{i\theta})$ is the g.l.b. of the spectrum of $W'(e^{i\theta})$, and by Lax (42), who showed that the literal generalization of the Szegö condition, $\log W' \in L_1(C)$, is inadequate.

Among the more distant work influenced by the contributions of Wiener and Kolmogorov we should mention (i) the extension of prediction theory to stationary random distributions due to K. Ito (30), Rosanov (61), and Balagangadharan (1); (ii) Cramer's exten-

sion of prediction theory to nonstationary processes in both discrete and continuous time (13, 14, 15). (iii) the firm initiation of a theory of linear least squares interpolation due to Yaglom and Rosanov (39, 67, 62).

All in all, the pioneering work of Wiener and Kolmogorov in prediction has stimulated much valuable mathematical activity in many lands.

19. The ramifications of prediction theory. It has become increasingly clear that linear prediction theory is a part of the general theory of one-parameter semi-groups of isometries on a Hilbert space. For instance, the Wold Decomposition of a weakly stationary, discrete parameter SP $(U^n f)_{n=-\infty}^{\infty}$ follows at once from the general equation, cf. Halmos (25),

$$(19.1) \quad \mathcal{H} = \bigcap_{n=1}^{\infty} V^n(\mathcal{H}) + \sum_{k=1}^{\infty} V^k(R^\perp)$$

which holds for any Hilbert space \mathcal{H} and any isometry V on \mathcal{H} onto $R \subseteq \mathcal{H}$. (Just take \mathcal{H} to be the present and past subspace \mathfrak{M}_0 of the given SP and V to be the restriction of U^* to \mathfrak{M}_0 .) The general theory of isometric semi-groups also embraces parts of functional analysis such as the theory of shifts, the Hardy class functions, etc. For instance, Beurling's Theorem (3, IV) on shift-invariant subspaces of the Hardy class H_2 also follows from (19.1) when \mathcal{H} and V are properly chosen.

Now many of the ideas and techniques suggested by prediction theory have proved useful in these related areas, and so the light emanating from Wiener's work on prediction has reached well beyond the confines of prediction theory proper. As examples of such distant ramifications of prediction theory we may mention (i) the deduction of the optimal-residual²⁸ factorization of a function f in the Hardy class H_2 from the Wold Decomposition of the SP $(e^{-kt}f)_{k=-\infty}^{\infty}$ (49); (ii) Lax's vectorial extension of Beurling's Theorem on shift-invariant subspaces using techniques suggested by q -variate prediction theory, (41); (iii) the canonical factorization of matrix-valued functions in the Hardy class H_2 obtained by the writer, again employing the ideas of q -variate theory, (49–53); (iv) the writer's enunciation and proof of the analogue of (19.1) for a continuous parameter semi-group $(S_t, t \geq 0)$ of isometries, in which he was guided by the situation in continuous time processes, (54); (v) the Hardy class theory on the torus due to Helson and Lowdenslager (27). Some of the recent work of Halmos and his collaborators, of Lowdenslager,

²⁸ Outer-inner, in Beurling's terminology.

Helson and others on the invariant subspace problem also falls into this category.

There are other regions of functional analysis where the viewpoint of prediction theory might fructify. An instance is the study of the operator identities of G. Baxter (2). Some of these identities were actually encountered in [186] in the derivation of the algorithm (17.6). Another instance is the theory of stationarity and prediction for individual time signals and certain associated algebras due to H. Furstenberg (22).

V. FILTER THEORY

20. Linear filters. Realizability. The only linear filters which Wiener considered at length were time-invariant ones which possess a differential weighting K in the time domain, i.e. filters for which the response g to an input signal f is given by

$$(20.1) \quad g(t) = \int_{-\infty}^{\infty} dK(t - \tau)f(\tau) = (dK*f)(t), \quad t \in (-\infty, \infty)$$

where K is a function of bounded variation on $(-\infty, \infty)$ and $*$ denotes convolution. An especially important subclass of such filters is that for which K is absolutely continuous on $(-\infty, \infty)$. Writing $K' = -W$, we then have

$$(20.2) \quad g(t) = \int_{-\infty}^{\infty} W(t - \tau)f(\tau)d\tau = (W*f)(t), \quad t \in (-\infty, \infty)$$

where $W \in L_1(-\infty, \infty)$. We shall refer to (20.1) and (20.2) as the $dK*$ and $W*$ filters.

To "know" the filters is to know the weightings K or W . A major problem of linear filter theory is to determine these weightings from a comparative analysis of the (observable) input and output signals f and g . Wiener made several contributions to this problem, cf. §§22, 23 below.

All physically realizable filters are *causal* (or *retrospective* or *non-anticipative*) in the sense that for each t

$$f_1 = f_2 \text{ on } (-\infty, t) \Rightarrow g_1 = g_2 \text{ on } (-\infty, t).$$

Obviously the filters $dK*$, $W*$ will be causal if and only if $K = 0 = W$ on $(-\infty, 0)$. In this case (20.1), (20.2) reduce to

$$(20.1') \quad g(t) = \int_{-\infty}^t dK(t - \tau)f(\tau), \quad t \in (-\infty, \infty),$$

$$(20.2') \quad g(t) = \int_{-\infty}^t W(t - \tau)f(\tau)d\tau, \quad t \in (-\infty, \infty).$$

One of Wiener's first contributions was to give the *spectral necessary and sufficient condition for causality of the filter W_** , viz.

$$(20.3) \quad \int_{-\infty}^{\infty} \{ |\log |\tilde{W}(\lambda)| | / (1 + \lambda^2) \} d\lambda < \infty,$$

where \tilde{W} is the indirect Fourier transform of W , cf. (1.1). This condition emerges from his joint work with Paley (§8), and engineers refer to it as the *Paley-Wiener condition*.

In case the filter W_* is causal, i.e. $W=0$ on $(-\infty, 0)$, \tilde{W} has a (unique) holomorphic extension to the upper half plane Δ_+ . This extension is in the Hardy class H_1 on Δ_+ , and so admits a canonical optimal-residual ("outer-inner") factorization

$$\tilde{W}(z) = \Phi(z)\Psi(z), \quad z \in \Delta_+$$

where

$$(20.4) \quad \Phi(z) = \exp \left\{ \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{\lambda z + 1}{\lambda - z} \frac{\log |\tilde{W}(\lambda)|}{1 + \lambda^2} d\lambda \right\}, \quad z \in \Delta_+$$

and Ψ is itself factorable into a Blaschke product, a factor, $\exp(i(\alpha + az))$ ($a \geq 0$, α real), and an integral akin to (20.4) but with respect to a purely singular and nonpositive measure over $(-\infty, \infty)$. In case $\Psi(z) = 1$, i.e. $\tilde{W} = \Phi$, the filter W_* is said to be of *minimum phase type*. The weighting for a filter of this type is easier to retrieve from input-output data than that of an arbitrary causal filter W_* , cf. §21.

In many cases of practical interest \tilde{W} is a rational function (cf. §15). A second major problem in filter theory is that of *design* or *synthesis*: given a rational function \tilde{W} satisfying (20.3), to synthesize from so-called lumped, passive, electrical elements (resistances, capacitances, inductances) the filter W_* . The problem of design is fairly well understood when the filter is linear. Wiener made some compelling suggestions as to its solution for nonlinear filters (§28).

21. Periodic and pulse inputs. To understand Wiener's work on filter inputs (§§22–25) we must review briefly the prior theory of purely periodic (in particular, sinusoidal) inputs and pulse inputs.

For the filter W_* we see at once that if $f(t) = e^{-i\lambda t}$, then

$$(21.1) \quad g(t) = \sqrt{(2\pi)} \tilde{W}(\lambda) e^{-i\lambda t} = \sqrt{(2\pi)} \tilde{W}(\lambda) f(t).$$

Thus the response to a sinusoidal input is a sinusoidal signal of the same frequency $-\lambda$ but with a complex amplitude $\sqrt{(2\pi)} \tilde{W}(\lambda)$. The function which gives the amplitude modification, viz. $\sqrt{(2\pi)} \tilde{W}$, is

called the *frequency response function* of the filter.²⁹ By successively plugging in sinusoidal inputs of different frequencies $-\lambda$ into the filter and observing the output amplitudes $\sqrt{(2\pi)\tilde{W}(\lambda)}$, we can get \tilde{W} and thence W .

Again for the filter $W*$, since $W \in L_1(-\infty, \infty)$, it follows that if f is in $L_2(-\infty, \infty)$, i.e., f is an energetic but powerless pulse, then $g = W*f$ is also in $L_2(-\infty, \infty)$; moreover,

$$(21.2) \quad \tilde{g}(\lambda) = \sqrt{(2\pi)\tilde{W}(\lambda)}\tilde{f}(\lambda).$$

Thus \tilde{W} can be retrieved from a knowledge of the Fourier-Plancherel transforms \tilde{f}, \tilde{g} of the input and output signal. Now cf. (1.1), (4.1),

$$(21.3) \quad \begin{aligned} |\tilde{f}(\lambda)|^2 &= \text{the energy density of } f \text{ at frequency } -\lambda, \\ |\tilde{g}(\lambda)|^2 &= \text{the energy density of } g \text{ at frequency } -\lambda, \end{aligned}$$

and if T_t is the translation operator, then

$$(21.4) \quad \int_{-\infty}^{\infty} e^{it\lambda} (T_t g, f) dt = \tilde{g}(\lambda)\bar{\tilde{f}}(\lambda) = \sqrt{(2\pi)\tilde{W}(\lambda)}|\tilde{f}(\lambda)|^2.$$

Since the functions f, g can be measured, the “output-input covariances” $(T_t g, f)$ for different lags t can be computed digitally, or analogically (by apparatus involving delay devices, square law rectifiers and integrators). Hence in principle we can find the cross output-input energy density (21.4). On dividing this by the input energy density (21.3) which again is computable from the observed signal f , we can find $\tilde{W}(\lambda)$ from (21.4).

In case the filter $W*$ is causal and of minimum phase type, the (unique) extension of \tilde{W} in the upper half plane Δ_+ is given (cf. (20.4)) by

$$(21.5) \quad \begin{aligned} \tilde{W}(z) &= \exp \left\{ \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{\lambda z + 1}{\lambda - z} \frac{\log |W(\lambda)|}{1 + \lambda^2} dx \right\}, \quad z \in \Delta_+ \\ &= \exp \left[\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\lambda z + 1}{\lambda - z} \frac{1}{1 + \lambda^2} \log \left\{ \frac{1}{2\pi} \frac{|\tilde{g}(\lambda)|^2}{|f(\lambda)|^2} \right\} d\lambda \right], \\ &\quad z \in \Delta_+, \text{ by (21.2)}. \end{aligned}$$

Thus in principle \tilde{W} can be found merely from a knowledge of the input and output energy densities (21.3) without recourse to the cross

²⁹ More specifically, for an electric filter it is called the *impedance* or *admittance* or *voltage ratio*, according as $f=i$ & $g=v$, or $f=v$ & $g=i$, or $f=v$ & $g=v$, where i is the *current* and v is the *voltage*. We have defined this function to be $\sqrt{(2\pi)\tilde{W}}$ (and not $\sqrt{(2\pi)\tilde{W}}$) in order that it be holomorphic in the upper (and not lower) half plane for a causal filter.

energy density (21.4). This is one of the merits of minimum phase filters.

A fictitious, limiting case of a pulse f as the support dwindle to a single point and the energy becomes infinite is the Dirac δ "function" supported at $\{0\}$ and such that $\delta(0) = \infty$. One assumes that $\int_{-\infty}^{\infty} |\delta(t)| dt = 1$, $\int_{-\infty}^{\infty} |\delta(t)|^2 dt = \infty$, so that δ is an *impact of "unit impulse" but infinite energy*. Formally, the response of the filter $W*$ to input δ is seen to be W :

$$(21.6) \quad g(i) = (W * \delta)(i) = W(i), \quad i \in (-\infty, \infty).$$

g will have finite energy if $W \in L_2(-\infty, \infty)$, i.e. $W \in L_1(-\infty, \infty) \cap L_2(-\infty, \infty)$. In this case

$$(21.7) \quad \text{the energy density of } g \text{ at frequency } -\lambda = |W(\lambda)|^2.$$

The practical importance of (21.6) and (21.7) stems from the fact that as f approaches δ , g approaches W . Since approximations to δ are physically realizable by means of sharp, almost instantaneous impacts, we can find good approximations to W by subjecting our filter to such impacts and measuring its response.

22. Inputs of the class \mathcal{S} . One of Wiener's main contributions to linear filter theory was to realize the importance of admitting as inputs arbitrary signals of the class \mathcal{S} (§2) and to extend the theory of L_2 - and periodic signals (§21) to cover this more general case. His main theorem is that if

$$(22.1) \quad tW(t) \in L_1(-\infty, \infty), \quad (1 + |t|)W(t) \in L_2(-\infty, \infty),^{30}$$

then the response of the filter $W*$ to a signal f in \mathcal{S} is a signal $g \in \mathcal{S}'$, and the spectral distributions S_f , S_g of f , g are related by

$$(22.2) \quad S_g(\lambda) = 2\pi \int_{-\infty}^{\lambda} |W(u)|^2 dS_f(u), \quad \lambda \in (-\infty, \infty)$$

[FI, Lemma 29₆, p. 173; Theorem 30, p. 178].

By a straightforward analysis Wiener showed that $g \in \mathcal{S}$ when $f \in \mathcal{S}$, but he had to appeal to the condition (2.8) of §2 on the "quadratic variation of s " to prove that g is actually in \mathcal{S}' . Latent in his work are results on the output-input covariance and spectral distribution ϕ_{gf} , S_{gf} :

$$(22.3) \quad \begin{aligned} \phi_{gf}(t) &= (W * \phi_f)(t), \quad t \in (-\infty, \infty), \\ S_{gf}(\lambda) &= \sqrt{(2\pi)} \int_{-\infty}^{\lambda} \tilde{W}(u) dS_f(u), \quad \lambda \in (-\infty, \infty). \end{aligned}$$

³⁰ From which it follows readily that W and the two functions just mentioned are in fact in $L_1(-\infty, \infty) \cap L_2(-\infty, \infty)$.

Let the input signal f have an absolutely continuous spectral distribution S_f . It then follows that the spectra S_{of} , S_o are also absolutely continuous, and (22.3), (22.2) reduce to

$$(22.4) \quad S'_{of}(\lambda) = \sqrt{(2\pi)} \tilde{W}(\lambda) S'_f(\lambda), \quad S'_o(\lambda) = 2\pi |\tilde{W}(\lambda)|^2 S'_f(\lambda).$$

The first of these equations corresponds to (21.2). Corresponding to (21.3) we now have, cf. (4.2) et seq.,

$$(22.5) \quad \begin{aligned} S'_f(\lambda)/\sqrt{(2\pi)} &= \text{the power density of } f \text{ at frequency } -\lambda. \\ S'_o(\lambda)/\sqrt{(2\pi)} &= \text{the power density of } g \text{ at frequency } -\lambda. \end{aligned}$$

As in the case of signals $f, g \in L_2(-\infty, \infty)$ (§21), we can find the output-input covariance $\phi_{of}(t)$ for different lags t from observed data either by digital computation or by analogical devices, and in principle obtain the output-input power density $S'_{of}(\lambda)$. On dividing this by the input power-density $S'_f(\lambda)$ we will get $\tilde{W}(\lambda)$, cf. (22.4). In case the filter W^* is causal and of minimum phase type, we can as in §21 retrieve \tilde{W} from a knowledge of the power densities (22.5) alone without recourse to the cross power density S'_{of} .

We see that Wiener was able to obtain for filters W^* a rather complete extension of the classical theory of L_2 -inputs to S -inputs with absolutely continuous spectra.

Corresponding to the fictitious δ signal of the L_2 -theory (§21) we now have the so-called unit *white noise signal* w characterized by the properties

$$\phi_w(t) = \delta(t), \quad S'_w(\lambda) = 1/\sqrt{(2\pi)}, \quad t, \lambda \in (-\infty, \infty).$$

Since $S_w(\infty) - S_w(-\infty) = \infty$, the signal w has infinite power. We may regard w as the fictitious limiting case of a signal f in S with a flat spectral density, the support of which swells to fill up the entire interval $(-\infty, \infty)$. Formally, the response g of the filter W^* to the input w satisfies

$$(22.6) \quad \phi_g(t) = W(t), \quad S'_{gw}(\lambda) = \tilde{W}(\lambda), \quad S'_g(\lambda) = \sqrt{(2\pi)} |\tilde{W}(\lambda)|^2.$$

On comparing this with (22.5) and (21.7) we can conclude that:

If W satisfies (22.1), the power density function of the response of the filter W^* to the unit white noise input w is equal to the energy density function of its response to the input δ , i.e. to an impact of unit impulse.

This is Wiener's result, being just a clearer rendition of his more cryptic formulation: "the response of a linear resonator to a unit

Brownian motion input has the same distribution of power in frequency that its response to a single instantaneous pulse will have as a distribution of energy in frequency" [TS, p. 50, GHA, p. 116]. We shall discuss its precise interpretation in §23.

23. Brownian motion inputs. The pseudo-concepts of the impact signal δ of infinite energy and of the white noise signal w of infinite power (§§21, 22) can be explicated in terms of "distributions" à la L. Schwartz. Integration by parts is an important tool in this explication. Such use of partial integration actually goes back to Wiener's early work on Brownian motion [29]. In [GHA, §13] Wiener adopted this method to convert pseudo-assertions concerning the response of the filter W^* to white noise into bona fide assertions about the response of the filter $-dW^*$ to a Brownian motion input.

To explain Wiener's approach, let us first see how the pseudo-assertion, (21.6) about the response of W^* to δ , can be reinterpreted as a result concerning the response of $-dW^*$ to the *Heaviside input* $\chi_{[0,\infty)}$. This input invokes from the filter $-dW^*$ the response

$$g(t) = - \int_{-\infty}^{\infty} d_{\tau} W(t - \tau) \chi_{[0,\infty)}(\tau) = - \int_0^{\infty} d_{\tau} W(t - \tau) = W(t).$$

Here we assume that W is of bounded variation on $(-\infty, \infty)$ and is in $L_2(-\infty, \infty)$, so that $W(-\infty) = 0$. Thus, *the response of the filter $-dW^*$ to the Heaviside input $\chi_{[0,\infty)}$ is W* . This statement would be the Wienerian explication of the pseudo-statement: the response of the filter W^* to δ is W .

Wiener handled the white noise input in a similar way. Let $\{x(t, \alpha), t \in (-\infty, \infty), \alpha \in [0, 1]\}$ be a (separable) BMSP. The response g_α of the filter $-dW^*$ to a Brownian motion $x(\cdot, \alpha)$ is given by

$$(23.1) \quad g_\alpha(t) = - \int_{-\infty}^{\infty} d_{\tau} W(t - \tau) \cdot x(\tau, \alpha) \quad t \in (-\infty, \infty).$$

To ensure the existence of the last integral Wiener assumed [GHA, p. 225] that

$$(23.2) \quad W(t) = \frac{K(t)}{\sqrt{1 + t^2}}, \quad K \text{ of bounded variation on } (-\infty, \infty).$$

Obviously $W \in L_2(-\infty, \infty)$ and is itself of bounded variation on $(-\infty, \infty)$. Since for almost all α , $x(\cdot, \alpha)$ is continuous on $(-\infty, \infty)$ and $|x(t, \alpha)| \leq 2\sqrt{(|t| \log |t|)}, |t|$ large, the integral in (23.1) exists for almost all α . Now we can show by partial integration that there exists a fixed set $N \subseteq [0, 1]$ of zero Lebesgue measure such that for

each $t \in (-\infty, \infty)$, there is a version $y(t, \cdot)$ of the stochastic integral

$$(23.3) \quad \int_{-\infty}^{\infty} W(t - \tau) dx(\tau, \cdot)$$

such that

$$(23.4) \quad y(t, \alpha) = g_\alpha(t), \quad \alpha \in [0, 1] - N, \quad t \in (-\infty, \infty).^{31}$$

Let us observe that the stochastic integral (23.3) is precisely the one defined in (5.4) above. Our present $y(t, \cdot)$ is just a well chosen member of the equivalence class $y(t, \cdot)$ of (5.4). It follows from (23.4), (22.5) and (5.5) that

$$(23.5) \quad \begin{aligned} &\text{for almost all } \alpha \text{ the response of the filter } -dW^* \text{ to the Brownian} \\ &\text{motions } x(\cdot, \alpha) \text{ belongs to } S' \text{ and has the (same) power density} \\ &\text{function } |\tilde{W}(-\cdot)|^2. \end{aligned}$$

With Wiener we may consider (23.5) as the explication of the pseudo-statement that $|\tilde{W}(-\cdot)|^2$ is the power-density function of the response of the filter W^* to the white noise input w . The pseudo statement (22.7) may likewise be given the precise rendering:

$$(23.6) \quad \begin{aligned} &\text{If } W \text{ satisfies (23.2), then the power density function of the} \\ &\text{response of the filter } -dW^* \text{ to almost all Brownian motions is} \\ &\text{equal to the energy density function of its response to the Heaviside} \\ &\text{input } \chi_{[0, \infty)}. \end{aligned}$$

24. On harmonic analysis and linearity. From the preceding account it is apparent that harmonic analysis, classical and generalized, is a valuable tool in the study of time-invariant, linear filters. What, however, is the precise connection between harmonic analysis and linearity? Wiener had a clear understanding of this question, and this fact was crucial in the evolution of his ideas on nonlinear filters. Unfortunately his writings on the question are somewhat vague [204, pp. viii-x, 51-53, NPRT, p. 90]. We shall therefore make some clarifying remarks before proceeding to his work on nonlinear filters.

Let G be any (additive) abelian Haar-measured group and \hat{G} be its character group. Let S_t be the *translation operator* induced by $t \in G$ on the space of complex-valued functions on G :

$$\{S_t(f)\}(x) = f(x + t), \quad t, x \in G.$$

Let T be any linear operator on some subspace, e.g. $L_2(G)$, of such

³¹ The proof of (23.4) involves some delicate questions concerning measurability and the behavior of mobile sets of zero-measure, which Wiener seems to have slurred over. The writer is grateful to Professor T. Hida for assistance in settling these questions.

functions into itself, which commutes with each S_t . Then each character α of G is an eigenfunction of T :

$$T(\alpha) = Y(\alpha) \cdot \alpha,$$

where $Y(\alpha) = \{ T(\alpha) \}(0)$. It follows that if f has a “Fourier expansion” in terms of the characters

$$f(x) = \sum_j c_j \alpha_j(x), \quad x \in G, \quad \alpha_j \in \hat{G},$$

then

$$\{ T(f) \}(x) = \sum_j c_j Y(\alpha_j) \cdot \alpha_j(x).$$

Hence $T(f)$ can be found, if we know the “Fourier coefficients” c_j of f and the “transfer function” (or “impedance”) Y of T . In short, T is completely determined by the function Y on \hat{G} , i.e. in filter terminology, by T ’s “response” to the “character inputs” α . “Harmonic analysis” thus appears as the natural tool for the study of linear operators which commute with the translations S_t .

In the case of time-invariant linear filters, G and therefore \hat{G} are of course the additive group of real numbers, and the harmonic analysis just alluded to becomes the classical harmonic analysis, or for robust, nonperiodic signals Wiener’s generalized harmonic analysis, the character-inputs being sinusoidals. The transfer function Y becomes the frequency response function, i.e. the Fourier transform of the weighting function (§§20, 21).

Next let T be a *nonlinear* operator on some subspace of complex-valued functions on G , again commuting with the translations S_t . Then in general T is not determined by its action on the characters of G . Consequently “character-inputs” (i.e. sinusoidals) do not play any intrinsic role, and harmonic analysis ceases to be especially significant. An important problem is to find the right substitutes: an appropriate probe, and an appropriate analysis of the response of the filter to such probes. With encouragement from Vannevar Bush, Wiener kept pondering on this question starting in the twenties.

25. The probe for a nonlinear filter. Among Wiener’s many statements on the question of a probe, perhaps the most succinct and lucid is that occurring in his posthumously published book [211, p. 34–]. He writes:

The output of a transducer excited by a given input message is a message that depends at the same time on the input message and on the transducer itself. Under the most usual circumstances, a transducer is a mode of transforming messages, and our attention is drawn to the output message as a transformation of the input message. However, there are circumstances, and these chiefly arise when the input message carries a minimum of information, when we may conceive the information of the out-

put message as arising chiefly from the transducer itself. No input message may be conceived as containing less information than the random flow of electrons constituting the shot effect. Thus the output of a transducer stimulated by a random shot effect may be conceived as a message embodying the action of the transducer.

As a matter of fact, it embodies the action of the transducer for any possible input message. This is owing to the fact that over a finite time, there is a finite (though small) possibility that the shot effect will simulate any possible message within any given finite degree of accuracy. . . . That is, if we know how a transducer will respond to a shot-effect input, we know ipso facto how it will respond to any input.

The belief that the nature of a filter can be found by studying its response to the random stimulation it receives from its environment—by psychoanalysis so-to-speak rather than by lobotomy—came rather early in Wiener's development. Thus in [GHA, p. 215] we read: “Imagine a resonator—say a sea-shell—struck by a purely chaotic sequence of acoustical impulses. It will yield a response which still has a statistical element in it, but in which the selective properties of the resonator will have accentuated certain frequencies at the expense of others.”

The shot effect of which Wiener speaks is to be regarded as the physical realization of a white noise signal $w(\cdot, \alpha)$ or more accurately of a Brownian incrementary signal $dx(\cdot, \alpha)$, (cf. §§22, 23). Our experiences with the Taylor expansion of ordinary (nonlinear) functions on \mathbb{R}^n suggest that the response of our filter to a white noise signal should be expressible as a linear combination of multiple stochastic integrals

$$(25.1) \quad \int_{-\infty}^t K(s) dx(s, \alpha), \quad \int_{-\infty}^t \int_{-\infty}^s K(s, \sigma) dx(s, \alpha) dx(\sigma, \alpha), \text{ etc.}$$

Wiener had already encountered such integrals in his researches on the *homogeneous chaos* [108], [128], the motivation for which lay in ergodic theory and statistical mechanics. This work, he now found, had a direct bearing on the problem of nonlinear filters, cf. §26.

26. Fourier-Hermite series of a function in $L_2[0, 1]$. Wiener came upon this subject in 1938 in his attempts to free ergodic theory from dependence on one-one point transformations [108]. In the course of this work he originated the idea of a *homogeneous chaos over \mathbb{R}^n* , i.e. a function F on $\mathfrak{G}_n \otimes [0, 1]$, (\mathfrak{G}_n = the family of Borel subsets of \mathbb{R}^n) such that for each $\alpha \in [0, 1]$, $F(\cdot, \alpha)$ is finitely additive on \mathfrak{G}_n , and for each $t \in \mathbb{R}^n$, each $S \in \mathcal{B}_1$, and each $B \in \mathfrak{G}_n$, the sets

$$\{\alpha: F(B + \{t\}, \alpha) \in S\}, \quad \{\alpha: F(B, \alpha) \in S\}$$

have the same Lebesgue measure [108, §2]. The current term for such an F is: *stationary random measure*.

The simplest example of a homogeneous chaos is the independently scattered Brownian measure over \mathfrak{R} , defined on $\mathfrak{B}_1 \otimes [0, 1]$ by extending

$$F((a, b], \alpha) = x(b, \alpha) - x(a, \alpha), \quad a < b,$$

where $\{x(t, \alpha), t \in \mathfrak{R}, \alpha \in [0, 1]\}$ is the BMSP. This F is called a 1-dimensional pure chaos [108, §6]. The multiple Wiener stochastic integrals yield derived chaoses; e.g. over \mathfrak{R}^n we have the chaos

$$F(B, \alpha) = \int_B \cdots \int K(t) dx(t_1, \alpha) \cdots dx(t_n, \alpha),$$

due to $K \in L_2(\mathfrak{R}^n)$; here $t = (t_1, \dots, t_n)$. Next, let

$$f(t, \alpha) = \int_{\mathfrak{R}^n} \cdots \int K(t - \tau) dx(\tau_1, \alpha) \cdots dx(\tau_n, \alpha)$$

and define

$$F(B, \alpha) = \int_B \cdots \int f(t, \alpha) dt_1 \cdots dt_n.$$

Then F is called an *nth-degree homogeneous polynomial chaos* [108, §9]. Such considerations led Wiener to the idea of a hierarchy of mutually orthogonal subspaces of $L_2[0, 1]$, [19, p. 37 et seq.]. The following version of his theorem is an adaptation of Kakutani's (33, Theorem 1):

THEOREM 1. *Let $\mathfrak{S}_0 = \mathfrak{R}$, and for $n \geq 1$, \mathfrak{S}_n be the subspace of symmetric functions in $L_2(\mathfrak{R}^n)$. Then for each $n \geq 0$, there is a linear isometry G_n on \mathfrak{S}_n into $L_2[0, 1]$ such that*

$$L_2[0, 1] = \sum_0^\infty G_n(\mathfrak{S}_n), \quad G_m(\mathfrak{S}_m) \perp G_n(\mathfrak{S}_n), \quad m \neq n,$$

$$U_t\{G_n(\mathfrak{S}_n)\} = G_n(\mathfrak{S}_n),$$

where U_t is the unitary flow on $L_2[0, 1]$ induced by the flow of the BMSP, more accurately of white noise, on $([0, 1], \text{Borel}, \text{Leb.})$, cf. (5.6).³²

The expressions for G_0 and G_1 are easily found:

$$(26.1) \quad \begin{aligned} \{G_0(c)\}(\alpha) &= c, \quad c \in \mathfrak{R}, \\ \{G_1(\phi)\}(\alpha) &= \int_{-\infty}^\infty \phi(t) dx(t, \alpha), \quad \phi \in L_2(-\infty, \infty). \end{aligned}$$

Those for G_2, G_3 etc. are best understood from the remarkable rela-

³² Our G_n differs from Wiener's in that his $G_n(K_n, \alpha)$ is our $\sqrt{(n!)}\{G_n(K_n)\}(\alpha)$, K_n being in $L_2(\mathfrak{R}^n)$. Our G_n is Kakutani's W_n^{-1} .

tionship which subsists between the G_n and the orthonormal polynomials H_n , $n \geq 0$, of the $(0, 1)$ normal distribution over \mathfrak{R} :

$$\int_{-\infty}^{\infty} H_m(u) H_n(u) \frac{1}{\sqrt{(2\pi)}} e^{-u^2/2} du = \delta_{mn}.$$

The H_n are of course the *Hermite polynomials*

$$H_n(u) = (n!)^{-1/2} (-1)^n e^{u^2/2} D_u^n(e^{-u^2/2}), \quad u \in \mathfrak{R}.$$

The relationship in question is that if $\{\phi_n, n \geq 0\}$ is any o.n. subset of $L_2(\mathfrak{R})$, and p_0, \dots, p_n are non-negative integers, then [NPRT, p. 94]³³

$$\left(\left(\sum_{i=0}^m p_i \right)! \right)^{1/2} G_{\sum_0^m p_i} \left\{ X \phi_i^{(p_i)} \right\} = \prod_{i=0}^m (p_i!)^{1/2} H_{p_i} \{ G_1(\phi_i) \} \in L_2[0, 1].$$

In view of this and Equation (26.1), we may recast Theorem 1 as a result on orthonormal expansions. The following enunciation of this is suggested by Cameron and Martin (8, Theorem 1).³⁴

THEOREM 2 (Fourier-Hermite expansion). *Let $f \in L_2[0, 1]$, $\{\phi_n, n \geq 0\}$ be an o.n. basis for $L_2(\mathfrak{R})$, and let for all non-negative integers p_0, \dots, p_m ,*

$$a_{p_0, \dots, p_m} = \left(f(\cdot), \prod_{j=0}^m H_{p_j} \left\{ \int_{-\infty}^{\infty} \phi_j(\tau) dx(\tau, \cdot) \right\} \right).$$

Then

$$f(\cdot) = \sum_{(p_0, \dots, p_m)} a_{p_0, \dots, p_m} \prod_{j=0}^m H_{p_j} \left\{ \int_{-\infty}^{\infty} \phi_j(\tau) dx(\tau, \cdot) \right\},$$

the convergence being in the $L_2[0, 1]$ topology.

The a_{p_0, \dots, p_m} are called the *Fourier-Hermite coefficients* of f , and the last series the *Fourier-Hermite series* of f relative to the basis

³³ For functions f_{p_0}, \dots, f_{p_m} on $\mathfrak{R}^{p_0}, \dots, \mathfrak{R}^{p_m}$, respectively, we define on $X_{i=0}^m f_{p_i}$ on $\mathfrak{R}^{\sum_0^m p_i}$ by

$$(X_{i=0}^m f_{p_i})(t_1, \dots, t_{\sum_0^m p_i}) = \prod_{j=0}^m f_{p_j} \{ t_{\sum_0^{j-1} p_i + 1}, \dots, t_{\sum_0^j p_i} \}.$$

We define $f^{(p)} = f x \cdots x f$ (p times).

³⁴ To get their Theorem 1 from our Theorem 2 we must, however, replace $t \in (-\infty, \infty)$ by $t \in [0, 1]$, and our $\alpha \in [0, 1]$ by the path $x(\cdot, \alpha) \in C_0[0, 1]$, where $C_0[0, 1]$ is the support of *Wiener measure*, i.e., the space of continuous functions $x(\cdot)$ on $[0, 1]$ such that $x(0) = 0$. It is outside our purview to report on the important role played by Wiener measure over $C_0[0, 1]$ in contemporary analysis. We refer to the work of Cameron and Martin, M. Kac, Feynman, K. Ito and others.

$\{\phi_n, n \geq 0\}$ of $L_2(\mathbb{R})$. This series plays an important role in Wiener's theory of nonlinear filters.

27. The response of a nonlinear filter to white noise. Let $g(\cdot, \alpha)$ be the response of our filter to the white noise signal $w(\cdot, \alpha)$, $\alpha \in [0, 1]$. We assume first that the filter is *time-invariant*, i.e., if g is the response to the input f , then the translate $g(\cdot + h)$ is the response to the input $f(\cdot + h)$. Taking $f(\cdot) = w(\cdot, \alpha)$, and recalling that

$$(27.1) \quad w(\tau + t, \alpha) = w(\tau, T_t \alpha),$$

where T_t is the flow of white noise over $([0, 1], \text{Borel}, \text{Leb.})$, cf. (5.6), it follows from the time-invariance of the filter that

$$(27.2) \quad g(\tau + t, \alpha) = g(\tau, T_t \alpha), \quad g(t, \alpha) = g(0, T_t \alpha).$$

We assume next that our filter is *stable* in the sense that

$$(27.3) \quad g(0, \cdot) \in L_2[0, 1],$$

(and hence $g(t, \cdot) \in L_2[0, 1]$). Wiener's remarks [NPRT, p. 89] as to how this condition restricts the filter are not clear to the writer. (The nexus between the ideas under consideration and the subject of nonlinear oscillations needs investigation.)

By (27.3) and §26, Theorem 2, $g(0, \cdot)$ has a Fourier-Hermite development relative to any o.n. basis $\{\phi_n, n \geq 0\}$ of $L_2(\mathbb{R})$. From (27.1) and (27.2) it easily follows that to get the Fourier-Hermite series for $g(t, \cdot)$ we have only to replace the functions $\phi_n(\cdot)$ by their translates $\phi_n(\cdot - t)$. Thus

$$(27.4) \quad g(t, \cdot) = \sum_{(p_0, \dots, p_m)} a_{p_0, \dots, p_m} \prod_{j=0}^m H_{p_j} \left\{ \int_{-\infty}^{\infty} \phi_j(\tau - t) dx(\tau, \cdot) \right\},$$

$$t \in \mathbb{R}.$$

Finally, we assume that our filter is *causal*, i.e., if g_1, g_2 are its responses to inputs f_1, f_2 , then for each t ,

$$f_1 = f_2 \text{ on } (-\infty, t) \Rightarrow g_1 = g_2 \text{ on } (-\infty, t).$$

It then follows that in (27.4) the upper terminus of the integrals must be t and not ∞ . In effect the functions ϕ_n must be supported on $(-\infty, 0]$, i.e., $\{\phi_n, n \geq 0\}$ must be an o.n. basis for $L_2(-\infty, 0]$. To conclude, *each time-invariant, stable, causal filter can be characterized in terms of any o.n. basis $\{\phi_n, n \geq 0\}$ of $L_2(-\infty, 0]$ by the system of coefficients a_{p_0, \dots, p_m} occurring in the Fourier-Hermite expansion (27.4) of its response at any given moment to white noise.*

By pursuing Wiener's heuristic idea quoted in §25 one should be able to prove that the coefficients a_{p_0, \dots, p_m} suffice to determine the

response of our filter to *any* reasonable input, not just white noise. One would then be able to assert that *to "know" the filter is to know the coefficients a_{p_0, \dots, p_m} relative to any o.n. basis of $L_2(-\infty, 0]$* . Such a proof does not seem to have been carried out, however, and hence it is not clear (to the writer at least) if Wiener's theory is equivalent to the others, which have been suggested, based on less random probes, cf. e.g. D. Gabor (23).

28. **Synthesis and analysis of nonlinear filters.** In the *synthesis problem* the function $g(\cdot, \alpha)$ is prescribed, and we have to design a filter ("white box") for which the response to white noise $w(\cdot, \alpha)$ is this $g(\cdot, \alpha)$. In the *analysis problem* we are given a "black box" the response of which to white noise is observable, and we are asked to determine its characterizing coefficients a_{p_0, \dots, p_m} relative to some o.n. basis of $L_2(-\infty, 0]$, cf. (27.4).

Wiener solved both problems. In his solution a crucial step was his selection of the functions $L_n(-\cdot)$, where L_n is the *n*th Laguerre function:

$$L_n(t) = e^{-t} t^{\frac{n}{2}} \sum_{k=0}^n (-1)^{k+1} 2^{k+1/2} (k!)^{-1} \binom{n}{k} t^k, \quad t \in [0, \infty), \quad n \geq 0,$$

for the o.n. basis for $L_2(-\infty, 0]$. Letting $\phi_j(t) = L_j(-t)$ and writing $w_\alpha(t)$ instead of $w(t, \alpha)$ for convenience, we see at once that

$$\int_{-\infty}^{\infty} \phi_j(\tau - t) dx(\tau, \alpha) = \int_{-\infty}^t L_j(t - \tau) w_\alpha(\tau) d\tau = (L_j * w_\alpha)(t);$$

i.e., the term on the left is the response at instant t of the *linear filter* L_j* to the white noise signal w_α , cf. (20.2), (20.2'). Wiener usually refers to $(L_j * w_\alpha)(t)$ as the (j th) "Laguerre coefficient of the past of the input." It follows from (27.4) that relative to this Laguerre basis

$$(28.1) \quad g(t, \alpha) = \sum_{(p_0, \dots, p_m)} a_{p_0, \dots, p_m} \prod_{j=0}^m H_{p_j} \{ (L_j * w_\alpha)(t) \},$$

and that *our filter ("black" or "white") can be characterized by the system of Laguerre-Hermite coefficients a_{p_0, \dots, p_m} .*

Now in 1928 Wiener and Y. W. Lee showed how the linear filter L_j* ($j \geq 0$) can be built from lumped, passive electrical elements—the so-called *Laguerre networks*. As Wiener saw and as we shall now indicate, this possibility along with (28.1) permits a theoretical solution of both the synthesis and analysis problems.

In the synthesis problem the Laguerre-Hermite coefficients a_{p_0, \dots, p_m} are given. The equation (28.1) shows that we can approxi-

mate to a filter with these coefficients to any desired degree (with the RMS error criterion) by suitably cascading a large number of Laguerre networks with square law rectifiers (multipliers), scale amplifiers and summing circuits. For the analysis problem we find from (28.1) with $t=0$, that

$$a_{p_0, \dots, p_m} = \int_0^1 g(0, \alpha) \cdot \prod_{j=0}^m H_{p_j} \{ (L_j * w_\alpha)(0) \} d\alpha.$$

Since the Brownian flow ($T_t, t \in \mathbb{R}$) is ergodic, it follows from the Ergodic Theorem and the relations

$$g(0, T_t \alpha) = g(t, \alpha), \quad w(0, T_t \alpha) = w(t, \alpha)$$

that for almost all α

$$a_{p_0, \dots, p_m} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau}^{\tau} g(t, \alpha) \prod_{j=0}^m H_{p_j} \{ (L_j * w_\alpha)(t) \} dt.$$

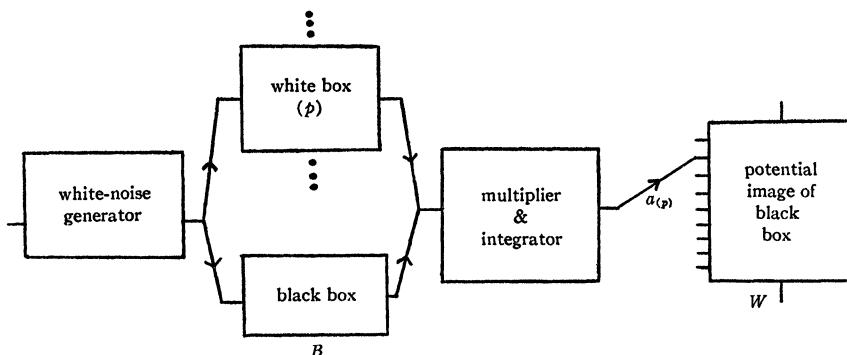
Now the second factor in the integrand is the response to the white noise signal w_α of a "white box," easily constructible from Laguerre networks, square law rectifiers and summing circuits, and the first factor $g(t, \alpha)$ is the "observed" response of our black box to this signal. Thus, by simultaneously bombarding our black box and the white box with the same white noise signal, and passing the outputs through a square law rectifier and integrating device we can approximate to the coefficients a_{p_0, \dots, p_m} . In this way both the synthesis and analysis problems can be tackled, in principle at least.

29. Reproduction, learning, self-organization. We can comment only very briefly on the bearing of Wiener's theory of filters on the larger questions of the reproducing, learning and self-organizing abilities of natural systems.

(i) Wiener realized that his solution of the analysis and synthesis problems for filters provided a theoretical³⁵ for the regeneration of filters. In the block diagram on p. 119 W is a white box comprising Laguerre networks, square law rectifiers, scale magnifiers and summing circuits, wired according to the scheme of (28.1). All parameters are fixed except those (inductances) which determine the coefficients a_{p_0, \dots, p_m} . W has a number of terminals into which signals controlling these parameters can be fed. Thus W is potentially capable of performing the operation of any time-invariant, stable, causal filter.

Now suppose that we wish to create an operative image of a black box B . We can obtain its (p_0, \dots, p_m) th Laguerre-Hermite coefficient by feeding the same white noise into B and the (known) (p_0, \dots, p_m) th white box, and averaging the product of the outputs (cf. §28, end). By feeding these averages into the coefficient-control-

³⁵ Professor R. B. Braithwaite's substitute for the overburdened term "model."



ling terminals of W (cf. Fig.), we can transform W into an actual operative image of B . Thus a machine can make another in its own likeness.

(ii) A *designed or purposive filter* is one whose response g to an input f approximates to $T(f)$, where T is some preassigned (well-defined) operator. The difference $g - T(f)$ between the actual and ideal response, or some norm thereof, is called its *error of performance*. A purposive filter is said to *learn*, if in the course of its operation the error of performance decreases. The existence of such filters is now well recognized.

Wiener visualized a learning filter as a system comprising a performing filter I coupled by feedback to a nonlinear filter II, [204, p. 173; 211, pp. 14, 20–21]. I carries out the routine of transforming input f into output g . II keeps a record of past inputs, outputs, and errors of performance of I, and has devices for the re-evaluation of the parameters governing this performance. Periodically, the system takes "time-out" to make this re-evaluation. The results are automatically fed back to I, and the routine resumed with improved efficiency. For instance, the component II of a learning anti-aircraft battery would record the long-time trajectories of incoming planes, and compute therefrom estimates of the covariances of the hypothetical underlying SP. For this it would have to include nonlinear devices such as square law rectifiers. When these improved estimates of the covariances are fed into component I of the battery, the new g 's produced by I will be closer to the ideal predictor $T(f)$ than before, i.e. the performance will improve.

Wiener compared this ability of a filter to fulfil its own purpose better by appropriately modifying its responses in the light of relevant external realities to *biological adaptation* (ontogenetic learning). He compared the propagation of such "self-educated" filters with racial or phylogenetic learning.

(iii) Wiener supplemented the ideas just recounted to explain the anti-entropic, self-organizing activity of other, less blue-printed,

natural systems. One such activity is the phenomenon of *gating* in the nervous system of animals [204, pp. 197–198]. This function, vitally important for the efficient use of nervous tissue, suggests the presence of a clock in the brain. How would such a clock survive the continual random bombardment it gets from its surroundings? Wiener's answer lay in his theory of *entrainment* or *attraction of frequencies*. The following is a brief outline of the mathematical side of this [NPRT, Lecture 8].

Consider an assembly of nonlinearly coupled oscillators, which in ideal isolation vibrate harmonically at the (same) fixed frequency ω . Because of the coupling the undisturbed oscillation $e^{i\omega t}$ gives way to

$$(29.1) \quad y(t, \alpha) = e^{i\omega t} f(t, \alpha)$$

where

$$(29.2) \quad f(t, \alpha) = \exp \left\{ i\epsilon \iint_{\mathbb{R}^2} K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) \right\}, \quad \epsilon > 0,$$

i.e., the “carrier” $e^{i\omega t}$ is “quadratically frequency-modulated by white noise.”³⁶ Actually for computational convenience Wiener took $\omega = 0$ in (29.1) (heterodyned frequency). He also assumed that K is symmetric and that ϵ is so small that in the expansion

$$f(t, \alpha) = 1 + i\epsilon \iint_{\mathbb{R}^2} K(t + \tau_1, t + \tau_2) dx(\tau_1, \alpha) dx(\tau_2, \alpha) + O(\epsilon^2)$$

the terms $O(\epsilon^2)$ can be neglected. He then showed that the spectral distribution F of the SP $y(\cdot, \cdot)$ has a sharp line at the (heterodyned) frequency 0, as well as an absolutely continuous part with density

$$(29.3) \quad F'(\lambda) = \frac{\epsilon^2}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} |Q(u, \lambda - u)|^2 du,$$

where Q is the direct Fourier-Plancherel transform of K in $L_2(\mathbb{R}^2)$. Actually the sharp line is smeared by the Doppler effect due to thermal noise, the resulting distribution being of Cauchy type [189]. The upshot is that the SP $y(\cdot, \cdot)$ has a spectral density which is the sum of (29.3) and a Cauchy density.

Now Wiener was guided by the case of an electric power generating system in which, through negative feedback, a number of alternators can maintain a sharp frequency despite variations of load. The fre-

³⁶ This is of course a simplification; we should also allow linear, cubic and higher order integrals in the { } in (29.2), and determine their contribution.

quencies "attract," i.e., the coupling is such that slow alternators are speeded up and fast ones slowed down. This led him to stipulate that

$$(29.4) \quad K(-t_1, -t_2) = -K(t_1, t_2)$$

—a kind of reaction principle. (29.4) entails that $Q(u, -u) = 0$ for all u ; whence $F'(0) = 0$. Putting the pieces together we find that the process $y(\cdot, \cdot)$ in (29.1) has a spectral density with a sharp peak at ω , with a dip and a small hill on either side. This profile is actually observed in the empirical spectrum of brain wave encephalographs.

Thus by dint of nonlinear couplings, the oscillators in an assembly can maintain narrow frequency bands, despite the presence of noise. Just as E. Hille saw semi-groups where the less initiated saw none, so Wiener detected entrainment in very diverse situations, e.g. the diurnal rhythm in many animals, the flashing of fire-flies in unison, the lumping in the periods of the asteroids, the breakdown of the earlier models of electra and comet airplanes. In [210], which was perhaps his last mathematical paper, Wiener attempted to show that a Hamiltonian or even low dissipative dynamical system, excited by random turbulence, would under certain circumstances generate nonlinear oscillations confined to narrow frequency bands. He even felt that quantum phenomena could be explained in this way. But the systematic discussion of Wiener's ideas on quantum theory, statistical mechanics and brain waves is beyond our scope.

30. Epilogue. Much remains to be done, especially at the nonlinear level, to set up Wiener's work on filters as a rigorous mathematical discipline. There are points of contact between his ideas and those conceived in the theories of nonlinear oscillations, of higher order spectra and general automata, which obviously need systematic exploration.

The ideas underlying Wiener's theory are very far reaching, and touch upon the very concepts of existence and progress. A purposive filter is able to produce a local zone of organization by absorbing energy and information from its environment, and maintaining homeostasis by feedback. But every filter must in the course of time wither away. As Wiener remarked, "The paradox of homeostasis is that it always breaks down in the end."³⁷ The value of existence should thus be gauged not in terms of sheer survival but in terms of the necessarily fugitive pursuit of anti-entropic activity. In [177, pp. 324–325] Wiener describes this predicament eloquently in relation to his own existence. He concludes: "The declaration of our own nature and the attempt to build up an enclave of organization in the

³⁷ Quotation from an unpublished manuscript.

face of nature's overwhelming tendency to disorder is an insolence against the gods and the iron necessity that they impose. Here lies tragedy, but here lies glory too."

But there is another aspect to this matter with which Wiener was also concerned. In the life of a highly organized filter may come moments when the fulfillment of its phylogenetic responsibilities might necessitate its own destruction. For phylogenetic survival such filters must learn a new art—that of sacrifice. The practice and theory of this art is religion, and mankind has always associated its highest manifestation with martyrdom. Can a theory of sacrifice be formulated within the framework of an enlarged theory of filters? Wiener mused on this and related questions in his Terry Lectures at Yale and in his posthumously published book [211]. Some of his ideas are profound, and suggest a way to free modern thought from the lurking inconsistency between the scientific and religious positions, which A. N. Whitehead and other great thinkers have found so vitiating.³⁸

In the course of his life Wiener tried to entrain many a filter to his point of view. His presence has been a great stimulus and a great challenge. One way we can express our affection and gratitude is to strive for an intellectual and moral climate in which others like him may arise. A more immediate task is to explore further the many ideas he has left behind. On the mathematical side, we have the problems of nonlinear innovations, of entrainment in relation to nonlinear oscillations, and the clarification of the notion of a stable filter, not to mention many others in areas like quantum mechanics, which we have not discussed.

FREQUENTLY QUOTED WORKS OF NORBERT WIENER

- GHA *Generalized harmonic analysis*, Acta Math. 55 (1930), 117–258.
 FI *The Fourier integral and certain of its applications*, Cambridge Univ. Press, New York, 1933.
 TS *Extrapolation, interpolation and smoothing of stationary time series with engineering applications*, The MIT Press, Cambridge, Mass.; Wiley, New York; Chapman & Hall, London, 1949.
 NPRT *Nonlinear problems in random theory*, The MIT Press, Cambridge, Mass. and Wiley, New York, 1958.

OTHER REFERENCES

1. K. Balagangadharan, *The prediction theory of stationary random distributions*, Mem. Coll. Sci. Univ. Kyoto Ser. A 33 (1960), 243–256.
2. G. Baxter, *An operator identity*, Pacific J. Math. 8 (1958), 649–663.

³⁸ A. N. Whitehead, *Science and the modern world*, Cambridge (1933), p. 94.

3. A. Beurling, *On two problems concerning linear transformations in Hilbert space*, Acta Math. 81 (1949), 239–255.
4. G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. 17 (1931), 656–660.
5. S. Bochner, *Lectures on Fourier Integrals*, Annals of Mathematics Studies No. 42, Princeton, N.J., 1959.
6. ———, *Harmonic analysis and the theory of probability*, Univ. of Calif. Press, Berkeley, Calif., 1955.
7. H. W. Bode and C. E. Shannon, *Linear least square smoothing and prediction theory*, Proc. Inst. Elec. Radio Engrs 38 (1950), 417–425.
8. R. H. Cameron and W. T. Martin, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. of Math. 48 (1947), 385–392.
9. J. Chover, *Conditions on the realization of prediction by measures*, Duke Math. J. 25 (1958), 305–310.
10. ———, *A theorem of integral transforms with an application to prediction theory*, J. Math. Mech. 8 (1959), 939–945.
11. H. Cramer, *On the representation of functions by certain Fourier integrals*, Trans. Amer. Math. Soc. 46 (1939), 191–201.
12. ———, *On the theory of stationary random processes*, Ann. of Math. 41 (1940), 215–230.
13. ———, *On the linear prediction problem for certain stochastic processes*, Ark. Mat. 4 (1959), 45–54.
14. ———, *On some classes of non-stationary stochastic processes*, Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press, Berkeley, Calif., 1960, 57–78.
15. ———, *On the structure of purely non-deterministic stochastic processes*, Ark. Mat. 4 (1961), 249–266.
16. S. Darlington, *Linear least-squares smoothing and prediction with applications*, Bell System Tech. J. 37 (1958), 1221–1294.
17. A. Devinatz, *The factorization of operator valued functions*, Ann. of Math. 73 (1961), 458–495.
18. C. L. Dolph and M. A. Woodbury, *On the relation between Green's functions and the covariances of certain stochastic processes and its application to unbiased linear prediction*, Trans. Amer. Math. Soc. 72 (1952), 519–550.
19. J. L. Doob, *Time series and harmonic analysis*, Proc. Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press, Berkeley, Calif., 1949, 303–343.
20. ———, *Stochastic processes*, Wiley, New York; Chapman & Hall, London; 1953.
21. E. B. Dynkin, *Theory of Markov processes*, Pergamon Press, New York, 1961.
22. H. Furstenberg, *Stationary processes and prediction theory*, Princeton University Press, Princeton, N. J., 1960.
23. D. Gabor, *Electronic inventions and their impact on civilization*, Imperial College of Science and Technology, London, March, 1959.
24. R. Gangolli, *Wide sense stationary sequence of distributions in Hilbert space and the factorization of operator valued functions*, J. Math. Mech. 12 (1963), 893–910.
25. P. Halmos, *Shifts of Hilbert spaces*, Crelle's Journal 208 (1961), 102–112.
26. O. Hanner, *Deterministic and non-deterministic stationary random processes*, Ark. Mat. 1 (1950), 161–177.
27. H. Helson and D. Lowdenslager, *Prediction theory and Fourier series in several variables*, Part I, Acta Math. 99 (1959) 165–202; Part II, 106 (1962), 175–213.

28. ———, *Vector-valued processes*, Proc. of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Univ. of Calif. Press, Berkeley, Calif. 1961, 203–212.
29. E. Hopf, *Ergoden Theorie*, Springer-Verlag, Berlin, 1931.
30. K. Ito, *Stationary random distributions*, Mem. Coll. of Sci. Univ. Kyoto Ser. A 28 (1953), 209–223.
31. Jang Ze-Pei, *The prediction theory of multivariate stationary processes. I*, Chinese Math. 13 (1963), 291–322.
32. S. Kakutani, Review of: *Extrapolation, interpolation, and smoothing of stationary time series*, by N. Wiener, Bull. Amer. Math. Soc. 56 (1950), 376–378.
33. ———, *Determination of the spectrum of the flow of a Brownian motion*, Proc. Nat. Acad. Sci. U.S.A. 36 (1950), 319–333.
34. R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, J. of Basic Engineering, March, 1961, 95–108.
35. K. Karhunen, *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Ann. Acad. Sci. Fenn. Ser. A I 37 (1947).
36. A. Y. Khinchin, *Korrelationstheorie der stationären stochastischen Prozesse*, Math. Ann. 109 (1934), 604–615.
37. A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, New York, 1950.
38. ———, *Curves in Hilbert space which are invariant with respect to a one-parameter group of motions*, Dokl. Akad. Nauk SSSR 26 (1940), 6–9.
39. ———, *Interpolation and extrapolation of stationary random sequences*, Izv. Akad. Nauk SSSR Ser. Mat. 5 (1941), 3–14.
40. ———, *Stationary sequences in Hilbert space*, Bull. Math. Univ. Moscow 2 (1941), no. 6.
41. P. Lax, *Translation invariant spaces*, Acta Math. 101 (1959), 163–178.
42. ———, *Translation invariant spaces*, Proc. of the International Symposium on Linear Spaces, Israel Acad. Sci. Jerusalem (1961), 299–306.
43. P. Levi, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937.
44. M. Loève, *Fonctions aléatoires du second ordre*, Note in P. Levi's Processus stochastiques et mouvement Brownien, Gauthier-Villars, Paris, 1948.
45. L. Mandel, *Fluctuations of light beams*, Progress of optics, Vol. II, New York, 1963, 183–248.
46. P. Masani, *The Laurent factorization of operator-valued functions*, Proc. London Math. Soc. 6 (1956), 59–69.
47. ———, *Cramer's Theorem on monotone matrix-valued functions and the Wold Decomposition*, Probability and Statistics, The Harald Cramer Volume, Wiley, New York, 1959, 175–189.
48. ———, *The prediction theory of multivariate stochastic processes. III*, Acta Math. 104 (1960), 141–162.
49. ———, *Shift invariant spaces and prediction theory*, Acta Math. 107 (1962), 275–290.
50. ———, *Sur la fonction génératrice d'un processus stochastique vectoriel*, C. R. Acad. Sci. Paris 249 (1959), 360–362.
51. ———, *Isomorphie entre les domaines temporel et spectral d'un processus vectoriel régulier*, C. R. Acad. Sci. Paris 249 (1959), 496–498.
52. ———, *Sur les fonctions matricielles de la classe de Hardy H_α* , C. R. Acad. Sci. Paris 249 (1959), 873–875; 906–907.

53. ———, *Une generalisation pour les fonctions matricielles de la classe de Hardy H_2 d'un theoreme de Nevanlinna*, C. R. Acad. Sci. Paris 251 (1960), 318–320.
54. ———, *On isometric flows on Hilbert space*, Bull. Amer. Math. Soc. 68 (1962), 624–632.
55. P. Masani and J. B. Robertson, *The time-domain analysis of continuous-parameter, weakly stationary stochastic processes*, Pacific J. Math. 12 (1962), 1361–1378.
56. R. F. Matveev, *On multidimensional regular stationary processes*, Theor. Probability Appl. 6 (1961), 149–165.
57. R. Nevanlinna, *Eindeutige analytische Funktionen*, (Zweite Auflage), Berlin, 1953.
58. J. B. Robertson, *Multivariate continuous parameter weakly stationary stochastic process*, Thesis, Indiana University, 1963.
59. M. Rosenberg, *The square-integrability of matrix-valued functions with respect to a non-negative hermitian measure*, Duke Math. J. 31 (1964), 291–298.
60. Iu. A. Rosanov, *The spectral theory of multi-dimensional stationary random processes with discrete time*, Uspehi Mat. Nauk, 13 (1958), 93–142. (Russian)
61. ———, *On the extrapolation of generalized stationary random processes*, Theory Probability Appl. 4 (1959), 465–472. (Russian)
62. ———, *Interpolation of stationary processes with discrete time*, Dokl. Acad. Nauk SSSR 130 (1960), 730–733=Soviet Math. Dokl. 1 (1960), 91–94.
- 62'. ———, *Stationary random processes*, Moscow, 1963. (Russian)
63. M. Rosenblatt, *Stationary processes as shifts of functions of independent random variables*, J. Math. Mech. 8 (1959), 665–681.
64. B. B. Rossi, *Optics*, Addison-Wesley, Reading, Mass., 1957.
65. P. Whittle, *The analysis of multiple stationary time series*, J. Roy. Statist. Soc. Ser. B 15 (1953), 125–139.
66. H. Wold, *A study in the analysis of stationary time series*, 2nd ed. Almqvist and Wiksell, Stockholm, 1954.
67. A. M. Yaglom, *On the problem of linear interpolation of stationary random sequences and processes*, Uspehi Mat. Nauk 4 (1949), 173. (Russian)
68. ———, *Extrapolation, interpolation and filtering of stationary random processes with rational spectral densities*, Trudy Moskov Mat. Obsc. 4 (1955), 333–374. (Russian)
69. ———, *Some classes of random fields in n-dimensional space, related to stationary random processes*, Theor. Probability Appl. 2 (1957), 273.
70. ———, *Effective solutions of linear approximation problems for multivariate stationary processes with a rational spectrum*, Theor. Probability Appl. 5 (1960), 239–264.
71. ———, *An introduction to the theory of stationary random functions*, Prentice-Hall, Englewood Cliffs, N. J., 1962.
72. L. A. Zadeh and J. R. Ragazzini, *An extension of Wiener's Theory of Prediction*, J. Appl. Phys. 21 (1950), 645–655.
73. V. Zasuhin, *On the theory of multidimensional stationary random processes*, Dokl. Akad. Nauk SSSR 33 (1941), 435–437. (Russian)