An Extension of Wiener's Theory of Prediction

LOTHI A. ZADEH AND JOHN R. RAGAZZINI
Department of Electrical Engineering, Columbia University, New York, New York
(Received November 14, 1949)

The theory of prediction described in this paper is essentially an extension of Wiener's theory. It differs from the latter in the following respects.
1. The signal (message) component of the given time series is assumed to consist of two parts, (a) a non-random function of time which is representable as a polynomial of degree not greater than a specified number \( n \) and about which no information other than \( n \) is available; and (b) a stationary random function of time which is described statistically by a given correlation function. (In Wiener's theory, the signal may not contain a non-random part except when such a part is a known function of time.)
2. The impulsive response of the predictor or, in other words, the weighting function used in the process of prediction is required to vanish outside of a specified time interval \( 0 \leq t \leq T \). (In Wiener's theory \( T \) is assumed to be infinite.)

The theory developed in this paper is applicable to a broader and more practical class of problems than that covered in Wiener's theory. As in Wiener's theory, the determination of the optimum predictor reduces to the solution of an integral equation which, however, is a modified form of the Wiener-Hopf equation.

A simple method of solution of the equation is developed. This method can also be applied with advantage to the solution of the particular case considered by Wiener. The use of the theory is illustrated by several examples of practical interest.

I. INTRODUCTION

Prediction—in the broad sense of the term—consists essentially of estimating the values of some function of time on the basis of a time series which may or may not contain random errors. For instance, a typical problem in prediction is as follows. Given a time series \( s(t) \) which is composed of a signal \( s(t) \) and a random disturbance (noise) \( N(t) \); provide an estimate of \( s(t+a) \), \( a \) being a positive constant, as a continuous function of time. More generally, the quantity to be estimated may be a functional of \( s(t) \) such as \( ds/dt \), \( \int s dt \), \( \int s^2 dt \), etc. In forming such estimates the mathematical operations that may be employed are usually limited by practical considerations. Thus, in most cases the operator furnishing the estimate must be linear and fixed in addition to the obvious requirement of being physically realizable. The physical counterpart of such an operator is what is commonly known as a predictor or an estimator.

It is evident that a function of time cannot be predicted intelligently unless sufficient a priori information is available about both the function and the errors. The nature of such information, as well as the characteristics of the signal and noise, can assume a variety of forms. Of these the more common ones have been investigated in recent years with the result that for certain classes of time series it is now possible to design predictors which make optimum use of the a priori information concerning the signal and the noise. Thus, when the given time series is stationary and the correlation functions of the signal and noise are known, one can use Wiener's theory\(^1\) to arrive at the specifications of the optimum predictor, that is, one minimizing the mean-square value of the prediction error. On the other hand, when, as is often the case in practice, the given time series is non-stationary, the available theories of prediction, notably Phillips and Weiss' theory,\(^2\) do not lead to the best possible predictor except for a narrow class of time series. It is possible, however, to extend Wiener's theory in many different directions thereby making it applicable to a wider class of problems than is covered by either Wiener's or Phillips and Weiss' theories in their present forms. One such extension is discussed in this paper. A feature of the extension is that the signal (message) is assumed to consist of a stationary component superimposed on a non-random function of time which is known to be representable as a polynomial of degree less than or equal to a specified number \( n \). It will also be shown that the general method developed for treating this problem can be applied with advantage to the solution of many cases of practical

---


interest as well as the particular case considered by Wiener.

II. FORMULATION

Consider a given time series $e(t)$ which is the sum of a function $s(t)$ (signal) and a stationary random disturbance $N(t)$. Let $s^*(t)$ be the quantity to be estimated and let $s(t)$ be related to $s(t)$ through a given linear operator $K(p)$, i.e.,

$$s^*(t) = K(p)s(t).$$  \hspace{1cm} (1)

$K(p)$ may be thought of as the system function of an ideal predictor, i.e., a predictor capable of perfect prediction in the absence of noise. In many cases, particularly those involving actual prediction, the operator $K(p)$ is not physically realizable so that the process of estimation cannot be carried out exactly even in the absence of random disturbances.

Frequently it will be convenient to use a different, though equivalent representation of Eq. (1), i.e.,

$$s^*(t) = \int_{-\infty}^{\infty} k(\tau)s(t-\tau) d\tau,$$  \hspace{1cm} (2)

where $\tau$ is the variable of integration and $k(t)$ represents the impulsive response of the ideal predictor. $K(p)$ shall be referred to as the ideal prediction operator. As a matter of convenience, the more common of the many possible forms which $K(p)$ and $k(t)$ can assume are given in Table I.

<table>
<thead>
<tr>
<th>$s^*(t)$</th>
<th>$s(t)$</th>
<th>$\delta(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s^*(t) = s(t)$</td>
<td>Present value of $s(t)$</td>
<td>1</td>
</tr>
<tr>
<td>$s^*(t) = s(t)$</td>
<td>Present value of $s(t)$</td>
<td>$\delta(t)$</td>
</tr>
<tr>
<td>$s^*(t) = s(t)$</td>
<td>Present value of $s(t)$</td>
<td>$\delta(t)$</td>
</tr>
<tr>
<td>$s^*(t) = s(t-\alpha)$</td>
<td>Past or future value of $s(t)$</td>
<td>$e^{-\alpha}$</td>
</tr>
</tbody>
</table>

Note: $\delta(t)$ denotes a unit impulse at $t = 0$, and $\delta^n(t)$ stands for the $n$th derivative of $\delta(t)$ with respect to $t$ (time).

(c) $M(t)$ and $N(t)$ have zero mean and are uncorrelated. This assumption is introduced only for the purpose of simplification and is not essential to the analysis. The condition expressed by (c) prevails in most practical cases.

Referring to Fig. 1, these inputs are shown being applied to the actual predictor whose system function is $H(p)$ and whose impulsive response is $W(t)$. The output of the predictor, $e_0(t)$, may be expressed in operational form

$$e_0(t) = H(p)e_1(t),$$  \hspace{1cm} (4)

or, alternatively, in the form of a superposition integral

$$e_0(t) = \int_0^t W(\tau)e_1(t-\tau) d\tau.$$  \hspace{1cm} (5)

An important characteristic of the actual predictor is the so-called prediction or estimation error $\epsilon$, which is defined as the difference between the output of the predictor and the quantity to be estimated, $s^*(t)$. Equation-wise this is:

$$\epsilon = e_0(t) - s^*(t).$$  \hspace{1cm} (6)

If there were no noise and if $K(p)$ were physically realizable there would be no prediction error and $H(p)$ would be identical with $K(p)$. This, of course, is the trivial case of the prediction problem. In what follows it will be assumed that either because of the presence of noise or physical unrealizability of $K(p)$, or both, $H(p)$ cannot be the same as $K(p)$.

The available a priori information about $s(t)$ and $N(t)$ is assumed to consist of $n_0$, $\psi_M(\tau)$, and $\psi_N(\tau)$. The problem is to specify the system function or the impulsive response of a predictor that would minimize in a certain sense the prediction error $\epsilon = e_0(t) - s^*(t)$. By analogy with Wiener's theory it will be postulated that the optimum predictor is the one in which: (a) the ensemble mean of $\epsilon$ is equal to zero (for all values of $t$), and (b) the ensemble variance of $\epsilon$ is a minimum. Denoting the ensemble average by the symbol $\langle \rangle_0$, these conditions read:

(a) $\langle \epsilon \rangle_0 = 0$ or, equivalently, $\langle e_0(t) \rangle_0 = \langle s^*(t) \rangle_0$, \hspace{1cm} (7)

(b) $\sigma^2 = \langle \epsilon^2 \rangle_0 = \text{minimum},$  \hspace{1cm} (8)
where \( \sigma^2 \), the ensemble variance of \( e_s \), is equal to the mean-square value of the prediction error. In what follows, conditions (a) and (b) will be used as the basis for the determination of the optimum predictor.

### III. Determination of the Impulsive Response of the Optimum Predictor

It will be recalled that the output of a predictor may be expressed in the form of a superposition integral

\[
e_s(t) = \int_0^\infty W(\tau)e_s(t-\tau)d\tau
\]  

(9)

where \( \tau \) is a dummy variable and \( W(t) \) represents the impulse response of the predictor. In practice it is usually found necessary to restrict the duration of sampling of the input time series to a finite constant \( T \), meaning in other words that \( W(t) \) must be zero outside the interval \( 0 \leq t \leq T \). To place this property in evidence Eq. (9) will be written in the following form:

\[
e_s(t) = \int_0^T W(\tau) e_s(t-\tau)d\tau.
\]

(10)

In the limiting case where the duration of sampling is infinite \( (T \to \infty) \) Eq. (10) becomes identical with Eq. (9).

By hypothesis,

\[
e_s(t) = P(t) + M(t) + N(t).
\]

(11)

Substituting Eq. (11) into Eq. (10) and making use of the identity

\[
P(t-\tau) = P(t) - \frac{\tau^2}{2} \frac{\partial^2 P(t)}{\partial \tau^2} + \ldots + (-1)^{n-1} \frac{\tau^n}{n!} \frac{\partial^n P(t)}{\partial \tau^n},
\]

(12)

it is found that \( e_s(t) \) may be expressed as:

\[
e_s(t) = \mu_0 P(t) - \mu_1 \dot{P}(t) + \frac{\mu_2}{2} \ddot{P}(t) + \ldots + (-1)^{n-1} \frac{\mu_n}{n!} \frac{\partial^n P(t)}{\partial \tau^n}
\]

\[+ \int_0^T W(\tau) M(t-\tau)d\tau + \int_0^T W(\tau) N(t-\tau)d\tau,
\]

(13)

where \( \mu_0, \mu_1, \mu_2, \ldots \), etc., designate the moments of \( W(t) \), i.e.,

\[
\mu_n = \int_0^T \tau^n W(\tau)d\tau, \quad n = 0, 1, 2, \ldots, n.
\]

(14)

Since \( M(t) \) and \( N(t) \) are stationary (with zero mean), it follows that the ensemble means of \( e_s(t) \) and \( s^*(t) \) depend only on the non-random component of the signal, i.e.,

\[
\langle e_s(t) \rangle_n = \int_0^T W(\tau) P(t-\tau)d\tau
\]

(15)

or

\[
\langle e_s(t) \rangle_n = \mu_0 P(t) - \mu_1 \dot{P}(t) + \frac{\mu_2}{2} \ddot{P}(t) + \ldots + \frac{(-1)^n}{n!} \frac{\partial^n P(t)}{\partial \tau^n},
\]

(16)

and

\[
\langle s^*(t) \rangle_n = (K(p) s(t))_n
\]

(17)

or

\[
\langle s^*(t) \rangle_n = K(p) P(t).
\]

(18)

Comparing Eqs. (16) and (18), condition (a) is reduced to

\[
K(p) P(t) = \mu_0 P(t) - \mu_1 \dot{P}(t) + \frac{\mu_2}{2} \ddot{P}(t) + \ldots + \frac{(-1)^n}{n!} \frac{\partial^n P(t)}{\partial \tau^n},
\]

(19)

Equation (19), being an identity, determines the values of \( \mu_0, \mu_1, \ldots, \mu_n \). In other words, the ideal prediction operator \( K(p) \) determines through Eq. (19) the first \( n+1 \) moments of the impulsive response of the optimum predictor.

As an illustration of the foregoing statement consider a case where the quantity to be estimated is the derivative of \( s(t) \), i.e., \( s^*(t) = \dot{s}(t) \). For this case Eq. (19) reduces to

\[
\dot{P}(t) = \mu_0 P(t) - \mu_1 \dot{P}(t) + \frac{\mu_2}{2} \ddot{P}(t) + \ldots + \frac{(-1)^n}{n!} \frac{\partial^n P(t)}{\partial \tau^n},
\]

(20)

and a term by term comparison of the left-hand and right-hand sides of Eq. (20) yields:

\[
\mu_0 = \int_0^T W(\tau)d\tau = 0
\]

\[
\mu_1 = \int_0^T \tau W(\tau)d\tau = -1
\]

\[
\mu_2 = \int_0^T \tau^2 W(\tau)d\tau = 0
\]

\[
\cdots
\]

\[
\mu_n = \int_0^T \tau^n W(\tau)d\tau = 0.
\]

(21)

These, therefore, are the \( n+1 \) constraints which the impulsive response of a derivative estimating network must satisfy.

As the second example consider a case where \( K(p) s(t) = s(t-\alpha) \), \( \alpha \) being a positive or negative constant. For
this case Eq. (19) reads
\[ P(t - \alpha) = \mu_0 P(t) - \mu_1 \dot{P}(t) + \frac{\mu_2}{2!} \ddot{P}(t) + \cdots + (-1)^n \frac{\mu_n}{n!} P^{(n)}(t). \]  
(22)

Rewriting \( P(t - \alpha) \) as
\[ P(t - \alpha) = P(t) - \alpha \dot{P}(t) + \frac{\alpha^2}{2!} \ddot{P}(t) + \cdots + (-1)^n \frac{\alpha^n}{n!} P^{(n)}(t), \]  
(23)

and making in Eq. (22) a term-by-term comparison of the coefficients of \( P(t) \), \( \dot{P}(t) \), etc., it is easily found that:

\[ \mu_0 = \int_0^T W(\tau) d\tau = 1 \]

\[ \mu_1 = \int_0^T \tau W(\tau) d\tau = \alpha \]  
(24)

\[ \cdots \]

\[ \mu_n = \int_0^T \tau^n W(\tau) d\tau = \alpha^n, \]

which thus represent the constraints imposed upon \( W(\tau) \) in case the quantity to be estimated is \( s(t - \alpha) \).

The problem that remains to be solved is that of minimizing \( \sigma^2 \). For this purpose it will be necessary to develop an explicit expression for \( \sigma^2 \) in terms of \( W(t) \) and the auto-correlation functions of the signal and noise. Assuming that condition (a) is satisfied, it follows from inspection of Eqs. (6), (13), and (19) that the prediction error is given by the expression

\[ \varepsilon = \int_0^T W(\tau) [M(t - \tau) + N(t - \tau)] d\tau - k(t) M(t) \]  
(25)

or equivalently

\[ \varepsilon = \int_0^T W(\tau) [M(t - \tau) + N(t - \tau)] d\tau - \int_0^{\infty} k(\tau) M(t - \tau) d\tau \]  
(26)

where \( k(t) \) is the impulsive response of the ideal predictor. The mean-square value of \( \varepsilon \) may be written as

\[ \sigma^2 = \langle \varepsilon^2 \rangle = \lim_{L \to \infty} \frac{1}{L} \int_0^L \varepsilon^2 d\tau \]  
(27)

\[ \sigma^2 = \lim_{L \to \infty} \frac{1}{L} \int_0^L d\tau \left[ \int_0^T W(\tau) [M(t - \tau) + N(t - \tau)] d\tau - \int_0^{\infty} k(\tau) M(t - \tau) d\tau \right]^2 \]  
(28)

A typical term of Eq. (28) such as

\[ \lim_{L \to \infty} \frac{1}{L} \int_0^L d\tau \left[ \int_0^T W(\tau) M(t - \tau) d\tau \right]^2 \]  
(29)

is expressible in the form of a triple integral

\[ \int_0^T \int_0^T d\tau_1 d\tau_2 W(\tau_1) W(\tau_2) \times \lim_{L \to \infty} \frac{1}{L} \int_0^L M(t - \tau_1) M(t - \tau_2) d\tau \]  
(30)

which in view of the definition of the auto-correlation function of \( M(t) \), i.e.,

\[ \psi_M(\tau) = \lim_{L \to \infty} \frac{1}{L} \int_0^L M(t) M(t - \tau) d\tau \]  
(31)

may be written as

\[ \int_0^T \int_0^T W(\tau_1) W(\tau_2) \psi_M(\tau_1 - \tau_2) \]  
(2)

Proceeding similarly in the case of other terms, Eq. (28) reduces finally to the following expression:

\[ \sigma^2 = \int_0^T \int_0^T W(\tau_1) W(\tau_2) \psi_M(\tau_1 - \tau_2) \]  
(25)

\[ + \psi_N(\tau_1 - \tau_2) \]  
(26)

\[ - 2 \int_0^{\infty} \int_0^T W(\tau_1) k(\tau_2) \psi_M(\tau_1 - \tau_2) d\tau_1 d\tau_2 \]  
(27)

\[ + \int_0^{\infty} \int_0^{\infty} k(\tau_1) k(\tau_2) \psi_M(\tau_1 - \tau_2) d\tau_1 d\tau_2 \]  
(28)

where, to recapitulate: \( \tau_1, \tau_2 \) = dummy variables; \( W(t) \) = impulsive response of the predictor; \( \psi_M(\tau) \) = auto-correlation function of \( M(t) \) (\( M(t) \) is the stationary part of the input signal); \( \psi_N(\tau) \) = auto-correlation function of \( N(t) \) (\( N(t) \) is the input noise); \( k(t) \) = impulsive response of the ideal predictor.

Returning to the problem of minimization of \( \sigma^2 \) it will be noted first that the last term in Eq. (33) is independent of \( W(t) \) and hence, insofar as minimization of \( \sigma^2 \) is concerned, need not be considered. Second, it will be recalled that \( W(t) \) is subject to the \( n+1 \) constraints
expressed by Eq. (14); therefore, the problem of minimizing $\sigma^2$ with respect to the class of $W(t)$'s satisfying Eq. (14) reduces essentially to an isoperimetric problem in the calculus of variations. Following the standard approach to such problems, one is led to minimizing the following expression:

$$I = \sigma^2 - 2\lambda_0\mu_0 - 2\lambda_1\mu_1 - \cdots - 2\lambda_n\mu_n$$  \hspace{1cm} (34)

or, more explicitly

$$I = \int_0^T W(\tau_1)d\tau_1 \left[ \int_0^T W(\tau_2)[\psi_M(\tau_1 - \tau_2) + \psi_N(\tau_1 - \tau_2)]d\tau_2 - 2\int_{-\infty}^{\infty} k(\tau_2)\psi_M(\tau_1 - \tau_2)d\tau_2 \right]$$

$$- 2\lambda_0 - 2\lambda_1\tau_1 - \cdots - 2\lambda_n\tau_n$$ \hspace{1cm} (35)

where the constants $\lambda_0, \lambda_1, \cdots, \lambda_n$, are the Lagrangian multipliers. Proceeding in the usual manner, that is, setting the variation of $I$ equal to zero, it is easily found that $I$ and hence $\sigma^2$ is a minimum provided $W(t)$ satisfies the following integral equation:

$$\int_0^T W(\tau)[\psi_M(t - \tau) + \psi_N(t - \tau)]d\tau = \lambda_0 + \lambda_1t + \cdots + \lambda_nt^n + \int_{-\infty}^{\infty} k(\tau)\psi_M(t - \tau)d\tau, \quad 0 \leq t \leq T. \hspace{1cm} (36)$$

This equation together with the $n+1$ constraints expressed by Eq. (14) provides the basis for the determination of the optimum predictor. It will be observed that in the particular case where $n=0, T=\infty$, and $k(t) = \delta(t+\omega)$ [with $\delta(t)$ standing, as usual, for a unit impulse at $t=0$], Eq. (36) reduces to

$$\int_0^\infty W(\tau)[\psi_M(t - \tau) + \psi_N(t - \tau)]d\tau = \psi_M(t + \omega), \quad t > 0 \hspace{1cm} (37)$$

which is essentially the integral equation of Wiener's theory. On the other hand, in the special case where $M(t) = 0$, Eq. (36) reduces to

$$\int_0^T W(\tau)\psi_N(t - \tau)d\tau = \lambda_0 + \lambda_1t + \cdots + \lambda_nt^n,$$

$$0 \leq t \leq T \hspace{1cm} (38)$$

which is the integral equation of Phillips and Weiss' theory. Thus, the integral equations of Wiener's, and Phillips and Weiss' theories are special cases of Eq. (36).

**IV. SOLUTION OF THE INTEGRAL EQUATION**

In the general case where $\psi_M(\tau)$ and $\psi_N(\tau)$ are prescribed but otherwise arbitrary auto-correlation func-
tions, the complicated nature of the integral equation makes it appear that the solution of Eq. (36) is a formidable problem. In reality, the problem is not as difficult as it may seem, for by using a procedure to be described, the general case can be reduced to a special case which has a simple solution.

Preliminary to the discussion of this procedure it will be expedient to introduce the spectral densities of $M(t)$, $N(t)$, and $M(t)+N(t)$. Denoting these by $S_M(\omega^2)$, $S_N(\omega^2)$, and $S(\omega^2)$, respectively, and recalling that the spectral density of a function is the inverse Fourier transform of its auto-correlation function, it follows that

$$S_M(\omega^2) = \int_{-\infty}^{\infty} \psi_M(\tau)e^{-i\omega\tau}d\tau$$ \hspace{1cm} (39)

$$S_N(\omega^2) = \int_{-\infty}^{\infty} \psi_N(\tau)e^{-i\omega\tau}d\tau$$ \hspace{1cm} (40)

and

$$S(\omega^2) = S_M(\omega^2) + S_N(\omega^2).$$ \hspace{1cm} (41)

Now the spectral density function $S(\omega^2)$ may be factored into the product of two conjugate factors

$$S(\omega^2) = G(j\omega)G(-j\omega)$$ \hspace{1cm} (42)

such that both $G(j\omega)$ and $1/G(j\omega)$ are analytic in the right half of the $j\omega$-plane. Usually $S(\omega^2)$ is assumed to be a rational function of $\omega^2$ of the form

$$S(\omega^2) = [A(\omega^2)/B(\omega^2)],$$ \hspace{1cm} (43)

where $A(\omega^2)$ and $B(\omega^2)$ are polynomials in $\omega^2$. For such cases the process of factorization is quite straightforward as can be seen from the following examples:

(a) $S(\omega^2) = \omega^2; \quad G(j\omega) = j\omega$.

(b) $S(\omega^2) = \frac{1}{\omega^2 + \omega_0^2}; \quad G(j\omega) = \frac{1}{j\omega + \omega_0}.$

(c) $S(\omega^2) = \frac{\omega^2 + a^2}{\omega^4 + b^2\omega^2 + c^2}; \quad G(j\omega) = \frac{j\omega + a}{(j\omega)^2 + j\omega(b^2 + 2c)^2 + c^2}.$

To summarize, a rational spectral density function may be written as

$$S(\omega^2) = |G(p)|^2_{\omega=jw} \hspace{1cm} (44)$$

where $G(p)$ is of the form:

$$G(p) = \frac{Q(p)}{p^m + b_1p^{m-1} + \cdots + b_mp} \hspace{1cm} (45)$$

and the polynomials $Q(p)$ and $R(p)$ do not have any zeros in the right half of the $p$-plane.
where $A(\omega^2)$ is the numerator of $S(\omega^2)$. It will be noted that $A(\omega^2)$ is a polynomial of the form

$$A(\omega^2) = \gamma_0 + \gamma_1 \omega^2 + \cdots + \gamma_m \omega^{2m},$$

(50)
and correspondingly the auto-correlation function of $M'(t) + N'(t)$ is

$$\psi_M'(\tau) + \psi_N'(\tau) = \gamma_0 \delta(\tau) - \gamma_1 \delta^{(2)}(\tau) + \cdots + (-1)^m \gamma_m \delta^{(2m)}(\tau),$$

(51)
where $\delta^{(n)}(\tau)$ represents the impulse function of $n$th order [i.e., the $n$th derivative of the unit impulse function $\delta(\tau)$].

In addition to $\psi_M'(\tau) + \psi_N'(\tau)$, a number of other quantities associated with the input to $N_2$ enter the integral equation satisfied by $W_2(t)$. The significance of each of these quantities, as well as their expressions, are as follows:

(a) $S_M'(\omega^2) = $ spectral density of $M'(t)$

$$S_M'(\omega^2) = S_M(\omega^2) |R(\omega)|^2,$$

(52)

(b) $\psi_M'(\tau) = $ auto-correlation function of $M'(t)$

$$\psi_M'(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_M(\omega^2) |R(\omega)|^2 e^{j\omega \tau} d\omega.$$  

(53)

(c) $k'(t) = $ ideal impulsive response for $N_2$

$$k'(t) = [1/R(\tau)]k(\tau).$$

(54)

In terms of these quantities the integral equation satisfied by $W_2(t)$ reads:

$$\int_0^t W_2(\tau) [\psi_M'(t-\tau) + \psi_N'(t-\tau)] d\tau = \lambda_0' + \lambda_1't + \cdots + \lambda_m't^m,$$

$$+ \int_{-\infty}^t k'(\tau) \psi_M'(t-\tau) d\tau, \quad t \geq 0.$$  

(55)

It will be noticed that in the case of $W_2(t)$ the upper limit of the integral is infinity, while in the case of $W(t)$ [cf. Eq. (36)] it is $T$. The explanation for this difference is that $W_2(t)$ need not vanish for $t > T$, even though $W(t)$ is required to do so. Thus in general, $W_2(t)$ will be piecewise analytic in the interval $0 < t < \infty$ as is illustrated in Fig. 3. Denoting the parts of $W_2(t)$ extending over the intervals $0 \leq t \leq T$ and $T < t < \infty$ by $U(t)$ and $V(t)$, respectively, the relation connecting $W(t)$ and $W_2(t)$ [cf. Eq. (46)] may be rewritten in the following form:

$$W(t) = R(\tau)U(t) + V(t),$$

(56a)
and

$$0 = R(\tau)V(t).$$

(56b)

These relations show that $W(t)$ is completely determined by the part of $W_2(t)$ which extends over the interval $0 \leq t \leq T$; the form of $W_2(t)$ outside this interval is irrelevant to the determination of $W(t)$.

---

1 The appendix of a report by Bode, Blackman, and Shannon, "Data smoothing and prediction in fire-control systems," Research and Development Board, Washington, D. C. (August, 1948), contains a brief exposition of a method which is similar in certain respects to the method described here.

2 It is tacitly assumed that $R(\tau)$ does not have a zero at the origin or, in other words, that $S(\omega^2)$ does not have a pole at zero frequency.
Returning to the integral equation (55), it will be noted that the range of integration \(0 \leq \tau < \infty\) may be divided into two parts, \(0 \leq \tau \leq T\) and \(T < \tau < \infty\), involving \(U(t)\) and \(V(t)\), respectively. Since \(V(t)\) is determined by Eq. (56b) to within a finite number of constants, the integral equation in question degenerates into an integral equation involving only \(U(t)\):

\[
\int_0^T U(\tau)[\psi_M'(t-\tau)+\psi_N'(t-\tau)]d\tau = \lambda_0' + \lambda_1't + \cdots
\]

\[
+ \lambda_n't^n + \int_0^\tau k(\tau)\psi_M'(t-\tau)d\tau, 0 \leq t \leq T. \tag{57}
\]

Upon substitution of Eqs. (52), (53), and (54), and performing minor simplifications, Eq. (57) reads

\[
\int_0^T U(\tau)[\gamma_0\delta(t-\tau)-\gamma_1\delta(t-\tau)+\cdots
\]

\[
+ (-1)^m\gamma_m\delta(t-\tau)]d\tau = \lambda_0' + \lambda_1't + \cdots + \lambda_n't^n
\]

\[
+ \frac{1}{2\pi} \int_0^\infty S_M(\omega^2)K(j\omega)R(-j\omega)e^{it\omega}d\omega. \tag{58}
\]

Making use of the identity

\[
\int_0^T U(\tau)\delta(t-\tau)d\tau = p^nU(t), \tag{59}
\]

Equation (58) may be rewritten as

\[
[\gamma_0-\gamma_1p^2+\cdots+(-1)^m\gamma_mp^{2m}]U(t)
\]

\[
= \lambda_0' + \lambda_1't + \cdots + \lambda_n't^n
\]

\[
+ \frac{1}{2\pi} \int_0^\infty S_M(\omega^2)K(j\omega)R(-j\omega)e^{it\omega}d\omega. \tag{60}
\]

Since in this equation the left-hand side operator is simply \(A(-p^n)\) [cf. Eq. (50)], the integral equation (57) finally reduces to the following differential equation:

\[
A(-p^n)U(t) = \lambda_0' + \lambda_1't + \cdots + \lambda_n't^n
\]

\[
+ \frac{1}{2\pi} \int_0^\infty S_M(\omega^2)K(j\omega)R(-j\omega)e^{it\omega}d\omega. \tag{61}
\]

The general solution of this equation is of the form:

\[
U(t) = A_0' + A_1't + \cdots + A_n't^n + B_1'\exp(\alpha_1t)
\]

\[
+ B_2'\exp(\alpha_2t) + \cdots + B_{2m}'\exp(\alpha_{2m}t)
\]

\[
+ \frac{1}{2\pi} \int_0^\infty S_M(\omega^2)K(j\omega)R(-j\omega)e^{it\omega}d\omega,
\]

where \(A_0', A_1', \cdots, A_n'\) and \(B_1', B_2', \cdots, B_{2m}'\) are as yet undetermined constants, and \(\alpha_1, \alpha_2, \cdots, \alpha_{2m}\) are the roots of the characteristic equation

\[
A(-p^n) = 0. \tag{63}
\]

In brief, Eq. (62) provides an explicit expression for \(U(t)\) involving \(2m+n+1\) undetermined constants. Availability of such an expression reduces the problem of determination of \(W(t)\) to a relatively routine matter which is discussed in the following section.

V. DERIVATION OF AN EXPPLICIT
EXPRESSION FOR \(W(t)\)

Recalling that \(W(t)\) is related to \(U(t)\) through the operational relation

\[
W(t) = R(p)U(t), \tag{56a}
\]

and substituting \(U(t)\) as given by Eq. (62) into Eq. (56a), it is readily found that in the most general case \(W(t)\) is given by the following expression:

\[
W(t) = \left[u(t) - u(t-T)\right]\left[A_0 + A_1t + \cdots + A_n't^n \right]
\]

\[
+ B_1\exp(\alpha_1t) + \cdots + B_{2m}\exp(\alpha_{2m}t)
\]

\[
+ \frac{1}{2\pi} R(p) \int_0^\infty S_M(\omega^2)K(j\omega)R(-j\omega)e^{it\omega}d\omega.
\]

\[
+ C_1\delta(t) + \cdots + C_{l-m}\delta(t-m-1) \tag{64}
\]

where the \(A's\), \(B's\), \(C's\), and \(D's\) are as yet undetermined constants, and the unit step functions \(u(t)\) and \(u(t-T)\) are used simply to indicate that \(W(t)\) is zero outside the interval \(0 \leq t \leq T\). The impulse functions contained in the expression for \(W(t)\) arise from operation by \(R(p)\) on the discontinuities of \(U(t)\) at \(t=0\) and \(t=T\). It will be observed that the order of these impulse functions does not exceed \(l-m-1\), which is one-half the order of the zero of \(S(\omega^2)\) at infinity minus one. This is due to the fact that the first \(m-1\) derivatives of \(U(t)\) vanish at \(t=0\) and \(t=T\). It is not difficult to verify that if this would not have been the case, the mean-square error at the output of \(N_s\) would be infinite.

Having obtained the general expression for \(W(t)\) in the form of Eq. (64), there remains the problem of determination of the \(2l+n+1\) unknown constants. These can be found in the following manner:

\[
\text{Fig. 3. Form of the impulsive response of } N_s.
\]
1. Substituting $W(t)$ as given by Eq. (64) into the integral equation (36) and requiring that the equation be satisfied identically, leads to $2l$ linear homogeneous equations in the $A$'s, $B$'s, $C$'s, and $D$'s.

2. Substituting $W(t)$ as given by Eq. (64) into the $n+1$ moment equations

$$\int_0^T \tau W(\tau) d\tau = \mu_n, \quad \nu = 0, 1, 2, \ldots, n$$

(65)
yields other $n+1$ linear equations. These $n+1$ equations, together with the $2l$ equations obtained in (1), provide a system of $2l+n+1$ linear equations in the unknown constants. Solution of this system yields the values of the $A$'s, $B$'s, $C$'s, and $D$'s and thus completes the process of determination of $W(t)$.

It should be remarked that in some cases it is advantageous to deal with the system function $H(p)$ of the predictor, rather than with its impulse response $W(t)$. In such cases one can use a transformed form of the integral equation (36) which is as follows:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} H(j\omega)S(\omega) e^{j\omega t} d\omega = \lambda_0 + \lambda_1 t + \cdots + \lambda_n t^n + \frac{1}{2\pi} \int_{-\infty}^{\infty} S_M(p^2)K(j\omega)R(-j\omega)e^{j\omega t} d\omega.$$  

(66)

Using Eq. (64), the solution of this equation may be written directly as

$$H(p) = \int_0^T (A_0 + A_1 t + \cdots + A_n t^n) e^{-pt} d\tau$$

$$+ \frac{B_1}{p + \alpha_1} + \frac{B_2}{p + \alpha_2} + \cdots + \frac{B_{2m}}{p + \alpha_{2m}} + \frac{B_{2m} \exp(\alpha_{2m} T)}{p + \alpha_{2m}} e^{-pt} \frac{1}{2\pi} R(p) \int_0^T dt e^{-pt}$$

$$\times \int_{-\infty}^{\infty} \frac{S_M(p^2)}{A(p^2)} K(j\omega) R(-j\omega) e^{j\omega t} d\omega + C_1 + C_2 p + \cdots + C_{2m} p^{2m-1} + (D_1 + \cdots + D_{2m} p^{2m-1}) e^{-pt}.$$  

(67)

The undetermined constants occurring in this expression are found in the same manner as in the case of $W(t)$, that is, $H(p)$ as given by Eq. (67) is substituted into the integral equation (66) and the resulting expression is treated as an identity. The $2l$ linear relations between $A_0, A_1, A_2, \ldots$, etc., which are obtained in this manner are adjoined to the $n+1$ relations resulting from Eq. (65); then the system of linear equations in the unknown constants is solved for $A_0, A_1, A_2, \ldots$, etc.

In order to facilitate application of the techniques described in the preceding sections, a summary of the procedure for the determination of $W(t)$ (or $H(p)$) is given in Section VI. Furthermore, actual use of the procedure is illustrated by a few practical examples at the end of the section.

### VI. SUMMARY OF THE PROCEDURE FOR DETERMINATION OF $W(t)$ AND $H(p)$

The complete expressions for $W(t)$ (the impulsive response of the optimum predictor) and $H(p)$ (the system function of the optimum predictor) are given by Eqs. (64) and (67). In order to avoid the necessity for reference to preceding sections, the meaning of all symbols appearing in these equations is given:

$u(t)$ = unit step function.

$T$ = duration of sampling (settling time).

$A_0, A_1, \ldots, A_n, B_1, B_2, \ldots, B_{2m}, C_1, C_2, \ldots, C_{2m}, D_1, D_2, \ldots, D_{2m}$ = undetermined constants.

$n$ = degree of the polynomial component of the input signal.

$S_M(\omega^2)$ = spectral density of $M(t)[M(0)]$ is the stationary part of the input signal.

$S(\omega^2)$ = spectral density of $M(t) + N(t)[N(0)]$ is the input noise.

$A(p^2)$ = numerator of $S(\omega^2)$.

$B(p^2)$ = denominator of $S(\omega^2)$.

$Q(j\omega)$ = a factor of $A(p^2)$ containing all the zeros in right half of the $j\omega$-plane.

$R(j\omega)$ = a factor of $B(p^2)$ containing all the zeros in left half of the $j\omega$-plane.

$2l$ = degree of $B(p^2)$.

$2m$ = degree of $A(p^2)$.

$\alpha_1, \alpha_2, \ldots, \alpha_{2m}$ = roots of the characteristic equation $A(-p^2) = 0$.

$\delta(t)$ = unit impulse function.

$\delta^{(i)}(t)$ = $i$th derivative of $\delta(t)$.

The undetermined constants occurring in the expression for $W(t)$ [and $H(p)$] can be found in the following manner.

1. $W(t)$ as given by Eq. (64) is substituted into the integral equation (36) and the resulting expression is treated as an identity. This gives $2l$ homogeneous linear equations in the unknown constants. Same equations can be obtained by substituting $H(p)$, as given by Eq. (67), into the integral equation (66).

2. $W(t)$ as given by Eq. (64) is substituted into the $n+1$ constraint equations

$$\int_0^T \tau W(\tau) d\tau = \mu_n, \quad \nu = 0, 1, \ldots, n.$$  

(14)

where the $\mu_n$ are determined by the choice of the prediction operator $K(p)$ [cf. Eq. (19)]. The resulting $n+1$ linear equations in the unknown constants are adjoined to the $2l$ equations obtained from (1). The set of $2l+n+1$ linear equations thus obtained is solved for the undetermined constants $A_0, A_1, \ldots, D_{2m}$. This concludes the process of determining $W(t)$ [or $H(p)$].
VII. ILLUSTRATIVE EXAMPLES

Example 1. Wiener's Theory

Wiener's theory is, in the main, a study of the particular case in which $P(t)=0$, $T=\infty$ and $K(\omega)=e^{-a\omega}$. For this case Eq. (64) gives

$$W(t) = u(t)\frac{1}{2\pi} \left\{ \frac{1}{R(\omega)} \times \int_{-\infty}^{\infty} \frac{S_M(\omega^2)}{A(\omega^2)} R(-j\omega) e^{j\omega t} \, d\omega \right. \left. + B_1 \exp(\alpha_1 t) + \cdots + B_m \exp(\alpha_m t) \right\}. \quad (68)$$

The exponential terms appearing in Eq. (68) may be made to vanish through a slight rearrangement of the factors in the first term of Eq. (68). The resulting expression for $W(t)$ is the same as that obtained by using Wiener's theory, namely,

$$W(t) = u(t)\frac{1}{2\pi} \frac{R(\omega)}{Q(\omega)} \times \int_{-\infty}^{\infty} \frac{S_M(\omega^2)}{Q(-j\omega)} R(-j\omega) e^{j\omega t} \, d\omega. \quad (69)$$

The rearrangement amounts, essentially, to choosing a particular solution of Eq. (61) which differs from the one chosen before by the exponential terms of Eq. (68). The same result may be achieved directly by choosing $H_1(\omega)$ [cf. Eq. (47)] as

$$H_1(\omega) = \frac{R(\omega)}{Q(\omega)}. \quad (70)$$

With this choice of $H_1(\omega)$ [in place of the one expressed by Eq. (47)] the various quantities entering Eq. (55) become:

$$S'(\omega^2) = 1, \quad S'(\omega^2) = \delta(\omega), \quad S_M(\omega^2) = S_M(\omega^2) R(j\omega)/Q(j\omega), \quad K'(\omega) = K(j\omega) Q(j\omega)/R(j\omega),$$

and hence the integral equation (55) reduces to

$$W_2(t) = u(t)\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_M(\omega^2)}{Q(-j\omega)} R(-j\omega) e^{j\omega t} \, d\omega; \quad (55a)$$

Eq. (69) then follows immediately from the relation connecting $W(t)$ and $W_2(t)$ [cf. Eq. (46)].

Example 2

The assumptions made here are as follows:

1. $M(t) = 0$.
2. $n = 1$.
3. $\psi_N(t) = e^{-a|t|}$; $S_N(\omega^2) = \frac{2a}{\omega^2 + a^2}$.

The choice of the prediction operator is left open.

Solution

For this case $A(-p^2) = 2a$, and hence $\alpha_1 = \cdots = \alpha_m = 0$. Also, $l = 1$, $M = 0$; hence Eq. (64) gives

$$W(t) = A_0 + A_1 t + C_1 \delta(t) + D_1 \delta(t - T), \quad (70)$$

and

$$H(p) = \frac{A_0}{p} \left( 1 - e^{-pT} \right) + \frac{A_1}{p} \left( e^{-pT} + T e^{-pT} \right) + C_1 + D_1 e^{-pT}. \quad (71)$$

Substituting $H(p)$ as given by Eq. (71) into the integral equation

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2a}{\omega^2 - p^2} H(p) e^{\omega t} d\omega = \lambda_0 + \lambda_1 t, \quad 0 \leq t \leq T, \quad (72)$$

and requiring that this equation be satisfied identically, leads to the following relations:

$$a A_0 - A_1 - a^2 C_1 = 0, \quad (73)$$

$$a A_0 + (aT + 1) A_1 - a^2 D_1 = 0. \quad (74)$$

Furthermore, substituting $W(t)$ as given by Eq. (70) into the constraint equations

$$\int_0^T W(t) d\tau = \mu_0 \tag{75}$$

and

$$\int_0^T \tau W(t) d\tau = \mu_1 \tag{76}$$

yields

$$A_0 T + A_1 (T^2/2) + C_1 + D_1 = \mu_0, \quad (77)$$

and

$$A_0 (T^2/2) + A_1 (T^3/3) + D_1 T = \mu_1. \quad (78)$$

The unknown constants $A_0$, $A_1$, $C_1$, and $D_1$ can be readily found from the solution of Eqs. (73), (74), (77), and (78). Thus,

$$A_0 = \mu_0 \left( \frac{4a(aT^2 + 3aT + 3)}{(a^2T^2 + 6aT + 12)(aT + 2)} \right) \tag{79}$$

$$A_1 = -\mu_0 \frac{6a^2}{a^2T^2 + 6aT + 12} + \mu_1 \frac{12a^2}{T(a^2T^2 + 6aT + 12)} \tag{80}$$

$$C_1 = \mu_0 \frac{2(aT^2 + 9aT + 12)}{(a^2T^2 + 6aT + 12)(aT + 2)} \tag{81}$$

$$- \mu_1 \frac{6(aT + 2)}{T(a^2T^2 + 6aT + 12)} \tag{81}$$

$$A_0 = \mu_0 \left( \frac{4a(aT^2 + 3aT + 3)}{(a^2T^2 + 6aT + 12)(aT + 2)} \right) \tag{79}$$

$$A_1 = -\mu_0 \frac{6a^2}{a^2T^2 + 6aT + 12} + \mu_1 \frac{12a^2}{T(a^2T^2 + 6aT + 12)} \tag{80}$$

$$C_1 = \mu_0 \frac{2(aT^2 + 9aT + 12)}{(a^2T^2 + 6aT + 12)(aT + 2)} \tag{81}$$

$$- \mu_1 \frac{6(aT + 2)}{T(a^2T^2 + 6aT + 12)} \tag{81}$$
On the other hand, in the case of the estimation
of \( s(t+\alpha) \) (i.e., the value of \( s(t) \) \( \alpha \)-seconds in the future)

\[
\mu_0 = 1 \\
\mu_1 = -\alpha.
\]  

[cf. Eq. (24)]

(86)

The shapes of \( W(t) \) for these two particular cases are illustrated in Fig. 4.

**Example 3**

The case to be considered here is the same as that treated in Example 2, except that the auto-correlation function of \( N(t) \) is assumed to be of the form

\[
\psi_N(\tau) = e^{-\alpha |\tau|} \cos \omega_0 \tau,
\]  

(87)

with the associated spectral density function being

\[
S_N(\omega) = \frac{2\alpha (\omega^2 + \omega_0^2 + \omega^2)}{\omega^4 + 2(\omega^2 - \omega_0^2) \omega^2 + (\omega^2 + \omega_0^2)^2}.
\]  

(88)

This form of spectral density function is of considerable practical importance since it provides a reasonably good approximation to many of the actual spectra encountered in practice.

**Solution**

By Eq. (64), the weighting function for this case is of the form

\[
W(t) = A_0 + A_1 t + B_1 e^{\delta t} + B_2 e^{-\delta t} + C_1 \delta(t - T), \quad 0 \leq t \leq T,
\]  

(89)

where \( \delta = (\alpha^2 + \omega_0^2)^{1/2} \). Substituting \( W(t) \) as given by Eq. (89) into Eq. (36), and requiring that Eq. (36) be satisfied by \( W(t) \) establishes four linear algebraic equations between the six constants \( A_0, A_1, B_1, B_2, C_1, D_1 \). These are:

\[
-2ab^2 A_0 + 2(\omega^2 - \omega_0^2) A_1 - b^3 B_1 + b^2 B_2 + C_1 = 0,
\]

\[
2\omega^2 b^2 A_0 - 4a \omega_0^2 A_1 + b^3 (b - a) B_1 + b^2 (a + b) B_2 = 0,
\]

\[
-2ab^2 A_0 + 2(\omega^2 - a^2 - a \omega_0^2) A_1 + b^3 (b - a) e^{\delta T} B_1 + 2b (\omega^2 - a^2 - a \omega_0^2) B_2 + D_1 = 0,
\]

\[
2\omega^2 b^2 A_0 + 2\omega_0^2 (2a + b^2 T) A_1 + b^3 (b + a) e^{\delta T} B_1 + b^3 (b - a) e^{-\delta T} B_2 = 0.
\]  

(90)

The constraints imposed by the prediction operator \( K(p) \) are given by Eq. (14); they are:

\[
\int_0^T W(t) dt = \mu_0,
\]  

(75)

and

\[
\int_0^T \tau W(t) dt = \mu_1.
\]  

(76)
The requirement that $W(t)$ must satisfy Eqs. (75) and (76) leads to two additional linear equations. These are:

$$2b T A_0 + b T^2 A_1 + 2(b T^2 - 1) B_1 - 2(e^{-b T} - 1) B_2 + 2b C_1 + 2b D_1 = \mu_0,$$

and

$$3b^3 T^2 A_0 + 2b^2 T^3 A_1 + b^2 [e^{b T}(b T - 1) + 1] B_1 - 6[e^{-b T}(b T + 1) - 1] B_2 + 6b^2 D_1 = \mu_1. \quad (91)$$

The coefficients $A_0$, $A_1$, $B_1$, $B_2$, $C_1$, $D_1$ of Eq. (89) are the solutions of the six linear equations (90) and (91).

The expression for the mean-square value of the prediction error can be conveniently expressed in terms of $A_0$ and $A_1$. Thus, making use of Eq. (83) it is readily found that

$$\sigma^2 = (2a/b^2)(\mu_0 A_0 + \mu_1 A_1). \quad (92)$$

This completes essentially the solution of the problem.

ACKNOWLEDGMENTS

The authors wish to thank Mr. H. T. Marcy of The M. W. Kellogg Company and Mr. H. Sherman of the Watson Laboratories, AMC, for their encouragement and support of this work.