

ON THE RELATION BETWEEN GREEN'S FUNCTIONS AND COVARIANCES OF CERTAIN STOCHASTIC PROCESSES AND ITS APPLICATION TO UNBIASED LINEAR PREDICTION⁽¹⁾

BY

C. L. DOLPH AND M. A. WOODBURY

INTRODUCTION

As a result of the stimulus of the work of Wiener [1]⁽²⁾ and Kolmogoroff [2] the theory of linear prediction has developed during the past decade. In this paper we shall develop a method of solution for a class of problems which are natural generalizations of these that have been treated by the so-called auto-correlation theory as developed by Phillips and Weiss [3] or Cunningham and Hynd [4].

A basic difficulty in a linear prediction theory employing continuous observations appears to be that of solving the integral equations of the first kind that arise from the minimization of the error variance. Since, as in the auto-correlation theory, the prediction is based on a finite past history, such tools as Wiener's generalized Fourier analysis are not available and a general method of solution of the integral equations has not yet been devised. Here we develop a method of solution that requires hypotheses which are satisfied in many important physical applications.

Moreover, as will be seen from Part I of the present paper, the kernel of our method is based on the intrinsically interesting relationship that exists between covariance functions of random processes generated by driving n th order linear differential equations by so-called "pure noise" and the Green's function of a suitably defined self-adjoint equation of order $2n$. For physically stable linear differential equations with constant coefficients, it will be shown (Corollary 1.1) that such covariance functions are in fact Green's functions of a suitably defined self-adjoint problem. For linear differential equations with variable coefficients this is no longer true but such covariance

Presented to the Society, September 1, 1949 under the title *Optimal linear prediction of stochastic processes whose covariances are Green's functions*; received by the editors February 18, 1951 and, in revised form, November 16, 1951.

⁽¹⁾ This work has been sponsored partly and jointly by the Air Force and by the Office of Naval Research under the respective contracts W33-038-ac-14222 (Willow Run Research Center) and N6 ONR232-1 with the Engineering Research Institute, University of Michigan.

The authors wish to express their thanks for assistance received from Professors D. A. Darling, T. H. Hildebrandt, and E. H. Rothe, and to extend their gratitude to Mr. F. L. Spitzer for his aid in the evaluation of the examples of Part III.

⁽²⁾ The numbers in brackets refer to the references at the end of the paper.

functions still have almost all properties possessed by Green's functions (Theorem 1.1) and the missing properties are of no consequence to our method. For this purpose we make use of the formulation due to Doob [5] of processes generated by such stochastic differential equations. First, one defines the so-called ξ -process as follows: For any finite set $t_1 < t_2 < \cdots < t_n$ of values of the real parameter t , the Gaussian random variables

$$\Delta \xi_i = \xi(t_{i+1}) - \xi(t_i), \quad i = 1, 2, \cdots, n-1,$$

are mutually independent and satisfy

$$(0.1) \quad E[\xi(t) - \xi(s)] = 0, \quad E[\xi(t) - \xi(s)]^2 = |t - s|,$$

for all t and s .

Second, were the derivative of $\xi(t)$ to exist it would imply that for any $t_1 < \cdots < t_n$ the random quantities $\xi(t_1), \cdots, \xi(t_n)$ would be mutually independent as in "pure noise." Since it does not exist, it is necessary to interpret the symbolic differential equation

$$(0.2) \quad \sum_{k=0}^n a_k(t) y^{(n-k)} = \xi'(t)$$

as defining a process having derivatives up to the $(n-1)$ th order which satisfies the equation

$$(0.3) \quad \sum_{k=0}^n \int_a^b f(t) a_k(t) dy^{(n-k)}(t) = \int_a^b f(t) d\xi(t)$$

with probability one for each continuous function $f(t)$ and each pair of numbers a and b . The integrals appearing in (0.3) are defined as the limit in the mean of the usual Stieltjes sum whenever $f(t)$ is continuous for $-\infty \leq a \leq t \leq b \leq \infty$ even though $\xi(t)$ is not of bounded variation in any finite interval. In particular, this type of integration commutes with the operation of taking expected values.

Part II of this paper is devoted to the solution of two Stieltjes integral equations of the first kind that are prototypes of those that occur in the linear prediction theory as it will be formulated in Part III. Our method of solution is based on a generalization of the well known [6] relationship that exists between a self-adjoint differential boundary value problem and an integral equation of the first kind whose kernel is the Green's function of this problem. Even when the covariance function is a Green's function some generalization is necessary since the usual equivalence is invalidated by the finite range of integration which a realistic prediction theory requires. In general, a left inverse exists but it is necessary to choose appropriate discontinuities at the ends of the interval of integration before an inverse to the integral operator can be found. In the simplest case we treat an integral equation analogous

to those of References [3] and [4] and obtain (Theorem 2.1) an explicit solution by this generalized equivalence. In the more complicated case we obtain (Theorem 2.2) a solution in terms of the resolvent kernel of an inhomogeneous Fredholm integral equation of the second kind.

Part III is devoted to the formulation of the linear prediction problem for processes of second order; i.e., $E[x(t)]$ and $E[x(t)]^2$ exist for all t . For such processes it is well known [7] that many results concerning them can be translated into the terminology of a Hilbert space of functions defined on a given probability space. The inner product in this space is given by $E[x(t)x(s)]$ and geometrically the second order processes are precisely those which are, for any fixed t , points in this Hilbert space. That is, they are paths in it. Moreover, as has been shown by Loève [7], one can construct a Gaussian process that has the same mean and covariance functions, so that, in a sense, Gaussian processes represent all second order processes.

Since we shall employ neither the hypothesis of stationarity nor of metric transitivity, it is necessary that the covariance function $E[x(t)x(s)]$ be known from a priori considerations. Now it is known [7] that if the covariance function of a second order process $x(t)$ possesses continuous derivatives of order $(2q+1)$ and a symmetric derivative of order $(2q+2)$, then the process $x(t)$ is derivable q times with probability one. Consequently, if the known past is taken to be the interval $[-0 \leq t \leq T+]$, we shall employ prediction operators of the form

$$(0.4) \quad x_p(t_f) = \sum_{k=0}^q \int_0^T x^{(k)}(t) d_t w_k(t, t_f)$$

where the number q is determined as above by the differentiable properties of the known covariance function $E[x(t)x(s)]$. The functions $w_k(t, t_f)$, $k=0, 1, \dots, q$, appearing in (0.4) are functions of bounded variation in t , and the integral that appears may be defined, as has been done by Loève [7] in the nonstationary case, as limits in the mean of the usual Riemann-Stieltjes sums. Again this type of integration commutes with the operation of taking expected values. For the prediction problems formulated in part III, Theorems (3.1) and (3.2) show that these operators, while not of the most general possible form, do in fact give the "best" linear prediction for $q=0, 1$. Furthermore, it will be clear from the proof of these theorems that corresponding results hold for any finite q . Operators of the above form are the natural generalizations of those of the auto-correlation theory and appear to be general enough for most applications.

The actual prediction is of course unknown until the $w_k(t, t_f)$ have been chosen by the criterion that defines the "best" linear prediction or that which requires the variance $E[x_p(t_f) - x(t_f)]^2$ to be a minimum. This requirement may be interpreted in the language of Hilbert space. Possible predictors, as points in this space, are constrained by definition to the closed linear manifold

determined by $x(t)$ for $0 \leq t \leq T$. It is natural to choose for the prediction that point in this manifold which is the projection of $x(t_f)$ on it. This point is precisely that value which minimizes the above variance as $x_p(t_f)$ ranges over this manifold. The explicit determination of it requires the solution of certain integral equations for the unknown weights $w_k(t, t_f)$. As in the auto-correlation theory the prediction is required to be unbiased so that the above minimization process must be carried out subject to the constraint that $E[x_p(t_f)] = E[x(t_f)]$.

More specifically, we treat the case in which $x(t) = x_a(t) + e(t)$, where $x_a(t)$ and $e(t)$ are second-order processes. If $0 \leq t_f \leq T$ this is a pure filter problem, while if $t_f > T$ it is a problem in prediction and filtering. If $x_a(t)$ is considered as the message, say in a communication circuit, $e(t)$ as random noise or error, and $x(t)$ as the observed signal, one may say that the problem of interest here is that of suppressing the "noise" $e(t)$ from the actual message $x_a(t)$.

Explicitly, we treat two cases of this last problem. In the simplest case (Corollary 3.1), as in the auto-correlation theory, $x_a(t)$ is a function known up to n parameters a_k , $k=1, 2, \dots, n$, and expressible in the form $x_a(t) = \sum_{k=1}^n a_k g_k(t)$ in terms of some family of known functions $\{g_k(t)\}$, $k=1, 2, \dots, n$, such as $\{t^k\}$ or a set of orthogonal functions. Theorem 3.1 considers the general filtering and prediction problem concerning the second-order stochastic process $x_a(t)$ whose mean can be expressed in the above form, that is, $E[x_a(t)] = \sum_{k=1}^n a_k g_k(t)$. Since $x_a(t)$ is considered physically as the message, it is assumed that it is at least as "smooth" as the noise; that is, it has at least as many derivatives as $e(t)$.

The solutions to these problems given in Part III are limited by the restriction that the covariance of the error process $e(t)$ must be related to a Green's function. As the theory of Part I indicates, this will happen naturally in many physical applications. In particular, it has long been standard practice [11] in communication engineering to generate a desired type of noise by driving a linear invariable network by "pure noise" so that our method is not only applicable but permits an investigation of the possible advantages of the use of variable networks in prediction theory.

Even though most applications will probably be limited to stationary error processes, the use of ensemble averages rather than time averages still has the advantage that it permits the treatment of mixed problems in which the message $x_a(t)$ may be nonstationary. A specific example of this type of problem is discussed in addition to the example previously treated in [3] and [4]. Even in the previously treated case our method provides considerable insight.

Because of the relatively unwieldy formulas that are necessary, Parts II and III are devoted mainly to the simplest application of the results of Part I. The modifications necessary in other cases, however, are indicated throughout (cf. Theorem 3.2). In particular, the number q which occurs in the

definition (0.4) of the prediction operator must always be chosen as one less than the order of the differential equation generating the error process in accordance with the interpretation of (0.2). That is, if the stochastic process has derivatives up to the $(n-1)$ th order, a priori, the information contained in them should be employed, in any prediction.

PART I. THE RELATIONSHIP BETWEEN CORRELATION
AND GREEN'S FUNCTIONS

It is well known [8] that the stochastic process generated by the Langevin equation of Brownian motion is a stationary Markovian process whose correlation function is $r(t-s) = e^{-\beta|t-s|}$. The Langevin equation may be written symbolically as $dy/dt + \beta y = \xi'(t)$ and interpreted like equation (0.2).

It is easy to verify that the above correlation function is a constant multiple of the Green's function of the following Sturm-Liouville system on the interval $-\infty \leq t \leq \infty$:

$$d^2y/dt^2 - \beta^2y = 0, \quad \lim_{t \rightarrow -\infty} y(t) = \lim_{t \rightarrow \infty} y(t) = 0.$$

The differential equation of this system is seen to be the result of letting the adjoint of the Langevin equation operate on that equation. It will be shown that very nearly this same relationship persists for all second-order processes generated in this way, even when the linear equations involve variable coefficients.

Before stating this relationship exactly, it will be convenient to obtain a representation for the solution of a certain initial value problem of ordinary linear differential equations in terms of a one-dimensional analogue of the Riemann function. Although such a concept has been almost certainly used before we know of no reference to it⁽³⁾. Fortunately the explicit derivation of those properties that we need is immediate enough to be indicated here.

Before defining this quantity it will be convenient to introduce the following notation. If a function $R(t, s)$ has a k th partial derivative with respect to t and s we shall set

$$\frac{\partial^k R(t, s)}{\partial s^k} = \frac{\partial^k R(t, s)}{\partial \alpha^k} \quad \text{and} \quad \frac{\partial^k R(t, s)}{\partial t^k} = \frac{\partial^k R(t, s)}{\partial \omega^k}$$

respectively so that the meaning of say $\partial^k R(t, t)/\partial \alpha^k$ will be clear.

In this notation we have the following:

DEFINITION 1.1. (a) *Let*

$$L_t(y) = \sum_{k=0}^n a_k(t) y^{(n-k)}(t)$$

⁽³⁾ In the usual treatment of the initial value problem (cf. Hurewicz [10]) a Green's function is used. In contrast to this the Riemann function has no singularities in its derivatives.

be an n th order linear differential operator defined for all t , $-\infty < t < \infty$, where $a_0(t) \neq 0$ and where the functions $a_k(t)$, $k=0, 1, \dots, n$, are of class $C^{(n-1)}$;

(b) Let

$$M_t(y) = \sum_{k=0}^n (-1)^k [a_k(t)y(t)]^{(n-k)}$$

be the differential operator formally adjoint to $L_t(y)$; then the one-dimensional Riemann function $R(t, s)$ associated with the initial value problem

(1.1)
$$L_s(y) = f(s),$$

(1.2)
$$y^{(k)}(t_0) = 0, \quad k = 0, 1, 2, \dots, (n - 1),$$

is the uniquely determined solution of the adjoint equation

(1.3)
$$M_s[R(t, s)] = 0$$

that satisfies the conditions

(1.4)
$$\frac{\partial^k R(t, t)}{\partial \omega^k} = \begin{cases} 0, & k = 0, 1, \dots, (n - 2), \\ (-1)^{n-1}/a_0(t), & k = (n - 1). \end{cases}$$

Similarly, the adjoint Riemann function, $R^*(t_0, s)$, is defined as the uniquely determined solution of

$$L_s[R^*(t_0, s)] = 0$$

that satisfies the conditions

$$\frac{\partial^k R^*(t_0, t_0)}{\partial \omega^k} = \begin{cases} 0, & k = 0, \dots, (n - 2), \\ 1/a_0(t), & k = (n - 1). \end{cases}$$

As in Riemann's case the basic properties of these functions follow from the Lagrange identity which may be written here as follows:

(1.5)
$$\int_{t_0}^t [zL_s(y) - yM_s(z)] ds = \sum_{j=0}^{n-1} y^{(j)} \left\{ \sum_{k=j}^{n-1} (-1)^{(k-j)} [a_{(n-k-1)}(s)z]^{(k-j)} \right\} \Big|_{s=t_0}^{s=t}.$$

We summarize these basic properties in

LEMMA 1.1. *The Riemann functions $R(t, s)$ and $R^*(t_0, s)$ satisfy the relations:*

(a)
$$R(t, t_0) = R^*(t_0, t),$$

(b)
$$L_t[R(t, t_0)] = 0,$$

(c)
$$M_{t_0}[R^*(t_0, t)] = 0,$$

$$(d) \quad \frac{\partial^k R(t, t)}{\partial \alpha^k} = \frac{\partial^k R^*(t, t)}{\partial \alpha^k} = 0, \quad k = 0, 1, \dots, (n-2),$$

$$(e) \quad \frac{\partial^{n-1} R(t, t)}{\partial \alpha^{n-1}} = 1/a_0(t),$$

$$(f) \quad \frac{\partial^{n-1} R^*(t, t)}{\partial \alpha^{n-1}} = (-1)^{n-1}/a_0(t).$$

The basic property (a) follows from the identity (1.5) if we set $z = R(t, s)$ and $y = R^*(t_0, s)$. The remaining properties then follow from this and the definitions of $R(t, s)$ and $R^*(t_0, s)$. We are now able to prove

LEMMA 1.2. *The solution of the initial value problem defined by (1.1) and (1.2) is given by*

$$(1.6) \quad y(t) = \int_{t_0}^t R(t, s)f(s)ds$$

where $R(t, s)$ is the Riemann function defined above.

The representation (1.6) follows immediately from the identity (1.5) if we set $z = R(t, s)$, replacing $L_s(y)$ by $f(s)$. Conversely, in view of (d) and (e),

$$y^{(k)}(t) = \int_{t_0}^t \frac{\partial^k R(t, s)}{\partial \alpha^k} f(s)ds, \quad k = 0, 1, \dots, (n-1),$$

while

$$\begin{aligned} y^{(n)}(t) &= \int_{t_0}^t \frac{\partial^n R(t, s)}{\partial \alpha^n} f(s)ds + \frac{\partial^{n-1} R(t, t)}{\partial \alpha^{n-1}} f(t) \\ &= \int_{t_0}^t \frac{\partial^n R(t, s)}{\partial \alpha^n} f(s)ds + f(t)/a_0(t). \end{aligned}$$

Multiplying $y^{(k)}(t)$ by $a_k(t)$, adding, and using (b) of Lemma 1.1, we see that (1.6) satisfies (1.1).

In terms of the Riemann function $R(t, s)$ we can define a stochastic process which is a solution of the equation $L_t(y) = \xi'$, and which specializes to that given by Doob [5] when $L_t(y)$ has constant coefficients and a characteristic equation whose roots possess negative real parts. In order to have a process which is defined for all t , $-\infty < t$, we choose $t_0 = -\infty$ in Definition 1.1, interpret $L_t(y) = \xi'$ in the manner indicated in the introduction, and use Lemma 1.2 to define this process by the formula

$$y(t) = \int_{-\infty}^t R(t, v)d\xi(v).$$

(Incidentally, as in Doob's case, it follows from this definition that $y^{(k)}(t)$ will exist for $k=0, 1, \dots, (n-1)$ while $y^{(n)}(t)$ will not exist since $\xi'(t)$ does not exist.)

The covariance function of this process is readily computed. By definition,

$$r(t, s) = E[y(t)y(s)] = \int_{-\infty}^t \int_{-\infty}^s R(t, v)R(s, u)E[d\xi(v)d\xi(u)].$$

In view of the characteristic property (0.1) of the ξ -process it follows that

$$(1.7) \quad \begin{aligned} r(t, s) &= \int_{-\infty}^s R(t, v)R(s, v)dv && \text{(for } t \geq s) \\ &= \int_{-\infty}^t R(t, v)R(s, v)dv && \text{(for } t \leq s). \end{aligned}$$

We are now able to state and prove the main result of this section.

THEOREM 1.1. *The covariance function $r(t, s)$ of the stochastic process*

$$y(t) = \int_{-\infty}^t R(t, v)d\xi(v)$$

generated by the differential equation

$$L_t(y) = \xi'$$

where $L_t(y)$ and $M_t(y)$ satisfy Definition 1.1 has the following properties:

- (i) *for $t \neq s$, it satisfies the obviously formally self-adjoint differential equation $M_t L_t(y) = 0$, in each of its variables t and s ;*
- (ii) *it and its first $(2n-2)$ th partial derivatives are continuous at $t=s$;*
- (iii) *at $t=s$ its $(2n-1)$ th derivative with respect to t has the jump appropriate to any Green's function of the equation $M_t L_t(y) = 0$. Explicitly⁽⁴⁾:*

$$\lim_{\epsilon \rightarrow 0} \left[\frac{\partial^{2n-1} r(t, s)}{\partial \alpha^{2n-1}} \right]_{t=s-\epsilon}^{t=s+\epsilon} = (-1)^n / a_0^2(s).$$

In order to see that $r(t, s)$ has the property (i) of Theorem 1.1 we note that for $t > s$

$$L_t[r(t, s)] = \int_{-\infty}^s L_t[R(t, v)]R(s, v)dv = 0$$

since (b) of Lemma 1.1 holds. On the other hand for $t < s$, we note that $r(t, s)$ as given by (1.7) is expressed in the form (1.6) so that it follows from Lemma 1.2 that $L_t[r(t, s)] = R(s, t)$. Using (1.3) this immediately implies that

⁽⁴⁾ The normalization used is that of Ince [6]. For odd n this agrees with that of Courant Hilbert [9] since the coefficient of $y^{(2n)}$ in $M_t L_t(y)$ is $-[a_0^2(t)]$ in this case.

$$M_t L_t[r(t, s)] = M_t[R(s, t)] = 0.$$

An entirely similar argument shows that $M_s L_s[r(t, s)] = 0$, for $t \neq s$.

Properties (ii) and (iii) follow by direct if somewhat tedious calculation. From (1.7) and Lemma 1.1 it follows that for $t > s$,

$$(1.8) \quad \frac{\partial^k r(t, s)}{\partial \alpha^k} = \int_{-\infty}^s \frac{\partial^k R(t, v)}{\partial \alpha^k} R(s, v) dv, \quad k = 1, 2, \dots, (2n - 1).$$

The fact that derivatives of the form $\partial^k R / \partial \alpha^k$ exist for $n - 1 < k \leq 2n - 1$ of course follows by induction from the assumption that the coefficients $a_k(t)$ are of class $C^{(n-1)}$ since (b) of Lemma 1.1 can be written in the form

$$a_0(t) \frac{\partial^n R(t, s)}{\partial \alpha^n} = \sum_{k=1}^n a_k(t) \frac{\partial^{n-k} R(t, s)}{\partial \alpha^{n-k}}.$$

On the other hand, for $t < s$ it follows from (d) and (e) of Lemma 1.1 that

$$\frac{\partial^k r(t, s)}{\partial \alpha^k} = \int_{-\infty}^t \frac{\partial^k R(t, v)}{\partial \alpha^k} R(s, v) dv$$

for $k = 0, 1, 2, \dots, (n - 1)$. For $k = n$ we obtain

$$(1.9) \quad \frac{\partial^n r(t, s)}{\partial \alpha^n} = \int_{-\infty}^t \frac{\partial^n R(t, v)}{\partial \alpha^n} R(s, v) dv + R(s, t) \frac{\partial^{n-1} R(t, t)}{\partial \alpha^{n-1}}$$

while for $n < k \leq 2n - 1$ it is clear that additional terms of the form

$$\frac{\partial^m}{\partial t^m} \left[R(s, t) \frac{\partial^p R(t, t)}{\partial \alpha^p} \right]$$

will occur with $m + p < 2n - 1$, $m = 0, 1, 2, \dots, n - 1$, $p = 0, 1, 2, \dots, 2n - 1$. However, in view of the conditions (1.4) satisfied by $\partial^m R(t, t) / \partial \omega^m$ it is clear that there will be only one contribution from all of these terms when $t = s$ and that this contribution will occur for $k = 2n - 1$ and will come from the $(n - 1)$ th derivative of the last term of (1.9). Thus

$$(1.10) \quad \begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{\partial^k r(s - \epsilon, s)}{\partial \alpha^k} &= \int_{-\infty}^s \frac{\partial^k R(s, v)}{\partial \alpha^k} R(s, v) dv \\ &+ \delta_k^{2n-1} \left[\frac{\partial^{n-1} R(t, t)}{\partial \alpha^{n-1}} \frac{\partial^{n-1} R(t, t)}{\partial \omega^{n-1}} \right] \\ &= \int_{-\infty}^s \frac{\partial^k R(s, v)}{\partial \alpha^k} R(s, v) dv + (-1)^{n-1} \delta_k^{2n-1} / a_0^2(t). \end{aligned}$$

The last step follows from (e) of Lemma 1.1 and (1.4). Here δ_k^{2n-1} is the usual Kronecker delta. By setting $t = s$ in (1.8) and subtracting (1.10) from it we

obtain (ii) and (iii) and thus complete the proof of Theorem 1.1.

Although the results of Theorem 1.1 are sufficient for the applications of Part III, we note the following:

COROLLARY 1.1. *If the differential operator $L_t(y)$ of Theorem 1.1 has constant coefficients and is such that all of the roots of its characteristic equation possess negative real parts, then the covariance function $r(t, s)$ defined in Theorem 1.1 is the Green's function of the self-adjoint system $M_t L_t(y) = 0$,*

$$\lim_{t \rightarrow -\infty} y^{(k)}(t) = \lim_{t \rightarrow \infty} y^{(k)}(t) = 0, \quad k = 0, 1, 2, \dots, (n - 1).$$

To show that the above boundary conditions are satisfied we recall that it is known [10] from the elementary theory of linear differential equations with constant coefficients that the representation (1.6) can be put in the form

$$\sum b_k \int_{-\infty}^t (t - s)^{q_k} e^{\lambda_k(t-s)} f(s) ds$$

for some constants b_k , where q_k denotes the multiplicity of the root λ_k of the characteristic equation of $L_t(y)$. From this it follows readily that if $p = \min(t, s)$, then

$$r(t, s) = \sum \sum b_k b_l e^{(\lambda_k + \lambda_l)(t+s)} \int_{-\infty}^p (t - v)^{q_k} (s - v)^{q_l} e^{-(\lambda_k + \lambda_l)v} dv,$$

from which it is readily seen that the above boundary conditions are satisfied. Moreover, $r(t, s)$ is seen to possess the symmetry property characteristic of a Green's function of a self-adjoint problem.

In order to see that $r(t, s)$ is not a Green's function in general we note that the solution of the initial value problem

$$y' + f'(t)y = \xi', \quad \lim_{t \rightarrow -\infty} y(t) = 0,$$

is given by $y(t) = e^{-f(t)} \int_{-\infty}^t e^{f(\sigma)} d\xi(\sigma)$. Thus

$$\begin{aligned} r(t, s) &= e^{-[f(t)+f(s)]} \int_{-\infty}^s e^{2f(\sigma)} d\sigma && \text{(for } t \geq s) \\ &= e^{-[f(t)+f(s)]} \int_{-\infty}^t e^{2f(\sigma)} d\sigma && \text{(for } t \leq s). \end{aligned}$$

For this example, $M_t L_t(y) = 0$ is

$$(1.11) \quad \left[\frac{-d}{dt} + f'(t) \right] \left[\frac{d}{dt} + f'(t) \right] y = 0,$$

and it is readily verified by direct calculation that $r(t, s)$ has all of the prop-

erties of Theorem 1.1. The Green's function of the differential equation (1.14) under the boundary conditions

$$\lim_{t \rightarrow -\infty} y(t) = \lim_{t \rightarrow \infty} y(t) = 0$$

may be calculated explicitly and is

$$G(t, s) = \frac{e^{-[f(t)+f(s)]} \int_t^\infty e^{2f(\sigma)} d\sigma \int_{-\infty}^s e^{2f(\sigma)} d\sigma}{\int_{-\infty}^\infty e^{2f(\sigma)} d\sigma},$$

for $t > s$. If one defines

$$\begin{aligned} m &= \int_{-\infty}^\infty e^{2f(\sigma)} d\sigma, \\ k(t) &= \int_t^\infty e^{2f(\sigma)} d\sigma, \\ \phi(t) &= e^{-f(t)}, \end{aligned}$$

one sees that $r(t, s)$ is related to $G(t, s)$ as follows:

$$\begin{aligned} G(t, s) &= r(t, s) \frac{k(t)}{m} && (t > s) \\ &= r(t, s) \frac{k(s)}{m} && (t < s). \end{aligned}$$

In particular, then, $r(t, s)$ is the form

$$\begin{aligned} r(t, s) &= \phi(t)\phi(s)k(t) && (t > s) \\ &= \phi(t)\phi(s)k(s) && (t < s). \end{aligned}$$

Later we shall need the trivial remark that functions $\phi(t)$ and $\phi(t)k(t)$ are linearly independent if $k(t)$ is not identically equal to a constant.

The simplest nontrivial example of the above is given by

$$L_t(y) = \frac{dy}{dt} + ty = \xi',$$

so that

$$\dot{M}_t L_t(y) = -\frac{d^2y}{dt^2} + (1+t^2)y = 0;$$

this latter expression is a special case of Hermite's equation.

PART II. SOLUTIONS FOR A CLASS OF INTEGRAL EQUATIONS
OCCURRING IN PREDICTION THEORY

In this part we shall construct a solution to the integral equations

$$(2.1) \quad f(t) = \int_0^T r(t, s) dw(s),$$

$$(2.2) \quad f(t) = \int_0^T [z(t, s) + r(t, s)] dw(s),$$

which are prototypes of those that will be derived as necessary and sufficient conditions for the solution of the problems of prediction and filtering that will be discussed in Part III. Although, strictly, this part is independent of the theory of stochastic processes, the entire discussion will be motivated by that theory. In particular, the hypotheses on $r(t, s)$ are precisely those that the covariance functions discussed in Part I have been shown to possess. The theory of stochastic processes further affects the discussion here in that it is limited to the problem of constructing a solution to these equations apart from the interesting questions concerning the uniqueness of their solutions or their most general solution. Neither of these last two questions affects the application of these equations to prediction.

Although the function $w(t)$ above will be assumed to be of bounded variation, the solution constructed here will be limited to functions of a particularly simple character. Thus, it will be shown sufficient to assume that $w(t)$ has no singular part and that it possesses a continuous derivative in the open interval $0 < t < T$ with possible jumps at the ends 0 and T . For equation (2.1) we have the following

THEOREM 2.1. *Let $r(t, s)$ be a symmetric function of the form*

$$\begin{aligned} r(t, s) &= \phi_1(t)\phi_2(s) && \text{(for } t \geq s) \\ &= \phi_1(s)\phi_2(t) && \text{(for } t \leq s), \end{aligned}$$

where $\phi_1(T) \neq 0$, $\phi_2(0) \neq 0$ and where the functions ϕ_1 and ϕ_2 are linearly independent; further let $r(t, s)$ satisfy a second-order self-adjoint linear differential equation $L_t(y) = [p(t)y']' + q(t)y = 0$ for $t \neq s$ and let this differential equation possess a Green's function over an interval containing $(0, T)$; assume also that^(A)

$$\lim_{\epsilon \rightarrow 0} \left[\frac{\partial r(s + \epsilon, s)}{\partial t} - \frac{\partial r(s - \epsilon, s)}{\partial t} \right] = - \frac{1}{p(s)},$$

and that $f(t)$ is integrable and possesses a continuous second derivative. Then the integral equation $f(t) = \int_0^T r(t, s) dw(s)$ has a solution given by

$$w'(t) = -L_t[f(t)],$$

$$\bar{w}(T) = -\frac{\dot{p}(T)}{\phi_1(T)} [\phi_1'(T)f(T) - \phi_1(T)f'(T)],$$

$$\bar{w}(0) = \frac{\dot{p}(0)}{\phi_2(0)} [\phi_2'(0)f(0) - \phi_2(0)f'(0)],$$

where

$$\bar{w}(t) = [w(t + 0) - w(t - 0)].$$

In order to prove this theorem we make the hypothesis mentioned above concerning the nature of $w(t)$. Under this hypothesis, (2.1) can be replaced by

$$(2.3) \quad f(t) = \int_0^T r(t, s)w'(s)ds + \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0).$$

Because $r(t, s)$ satisfies the differential equation $L_t(y) = 0$ for $t \neq s$, it follows at once that if a solution exists one must have $w'(t) = -L_t[f(t)]$ for $0 < t < T$. In order to guarantee that a solution of this form exists, it must be shown that $\bar{w}(T)$ and $\bar{w}(0)$ can be chosen so that the expression

$$S = \int_0^T r(t, s) \{ -L_s[f(s)] \} ds + \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0)$$

reduces to $f(t)$ identically. The first term of this expression can be transformed by the Lagrange identity

$$(2.4) \quad \int_0^T [zL_s(y) - yL_s(z)]ds = [p(y'z - z'y)]_{s=0}^{s=T},$$

so that S can be written as

$$S = -\int_0^T f(s)L_s[r(t, s)]ds + \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0)$$

$$- \left\{ \dot{p}(s) \left[f'(s)r(t, s) - \frac{\partial r}{\partial s}(t, s)f(s) \right] \right\}_{s=0}^{s=t}$$

$$- \left\{ \dot{p}(s) \left[f'(s)r(t, s) - \frac{\partial r}{\partial s}(t, s)f(s) \right] \right\}_{s=t}^{s=T}.$$

A repetition of the standard argument used for Green's functions shows that the first term vanishes, while the use of the properties assumed about $r(t, s)$ enables us to reduce the above to

$$S = f(t) + \phi_2(t) \{ -\dot{p}(T) [\phi_1'(T)\phi_1(T) - \phi_1(T)\phi_1'(T)] + \bar{w}(T)\phi_1(T) \}$$

$$+ \phi_1(t) \{ \dot{p}(0) [f'(0)\phi_2(0) - \phi_2'(0)f(0)] + \bar{w}(0)\phi_2(0) \}.$$

Because of the assumed linear independence of the functions ϕ_1 and ϕ_2 , S can reduce to $f(t)$ identically only if the coefficients of $\phi_1(t)$ and $\phi_2(t)$ vanish. This yields the values of $\bar{w}(T)$ and $\bar{w}(0)$ given in the theorem.

It is evident that the above technique would not be adequate if $r(t, s)$ were associated with a self-adjoint linear differential equation of order higher than the second. The next case of interest in prediction occurs when $r(t, s)$ is associated with a fourth-order equation. For this case it is necessary to replace (2.3) by

$$(2.5) \quad f(t) = \int_0^T r(t, s)w'(s)ds + \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0) \\ + \bar{q}(T) \left. \frac{\partial r}{\partial \omega}(t, s) \right|_{s=T} + \bar{q}(0) \left. \frac{\partial r}{\partial \omega}(t, s) \right|_{s=0},$$

so that once again the number of possible jumps $\bar{w}(T)$, $\bar{w}(0)$, $\bar{q}(T)$, and $\bar{q}(0)$ equals the order of the differential equation satisfied by $r(t, s)$. That such a replacement is possible for this case will be shown in Part III. Once this has been accomplished, the above technique can be used again to determine $w'(t)$ and the above discontinuities can be determined once again, since $r(t, s)$ will involve four linearly independent functions. The fact that $\bar{q}(T)$ and $\bar{q}(0)$ do not drop out when the differential operator is applied to (2.5) does not cause any difficulty, and can be treated as in the next theorem.

The above method can be extended to the integral equation (2.2) which is a prototype of one occurring in the general filter and prediction problem. Specifically, we have the following

THEOREM 2.2. *Let $r(t, s)$ and $f(t)$ satisfy the hypotheses of Theorem 2.1 and let $K(t, s) = L_s[z(t, s)]$, its resolvent kernel $M(t, s)$, and $\partial M(t, s)/\partial t$ exist. If*

$$h(t) = f(t) - \bar{w}(T)z(t, T) - \bar{w}(0)z(t, 0), \\ A(t) = z(t, T) + \int_0^T z(t, s)M(s, T)ds, \\ B(t) = z(t, 0) + \int_0^T z(t, s)M(s, 0)ds, \\ C(t) = - \left\{ L_t[f(t)] + \int_0^T M(t, s)L_s[f(s)]ds \right\}, \\ D(t) = - \left\{ f(t) + \int_0^T C(s)z(t, s)ds \right\}, \\ E(t) = A(t)\bar{w}(T) + B(t)\bar{w}(0) + D(t),$$

a solution of the integral equation

$$f(t) = \int_0^T [z(t, s) + r(t, s)]dw(s)$$

is given by

$$w'(t) = - \left\{ L_t[h(t)] + \int_0^T M(t, s)L_s[h(s)]ds \right\}$$

and by those values of $\bar{w}(0)$ and $\bar{w}(T)$ that satisfy the linear system

$$\begin{aligned} \bar{w}(T)\phi_1(T) + p(T)[\phi_1(T)E'(T) - \phi_1'(T)E(T)] &= 0, \\ \bar{w}(0)\phi_2(0) - p(0)[\phi_2(0)E'(0) - \phi_2'(0)E(0)] &= 0. \end{aligned}$$

In order to establish these results⁽⁵⁾ we again make use of the hypotheses concerning the nature of $w(t)$ and replace (2.2) by

$$(2.6) \quad h(t) = \int_0^T [z(t, s) + r(t, s)]w'(s)ds + \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0).$$

The application of the differential operator $L_t(g)$ associated with $r(t, s)$ reduces the last expression to this Fredholm integral equation of the second kind:

$$-L_t[h(t)] = w'(t) - \int_0^T K(t, s)w'(s)ds.$$

Since the resolvent kernel $M(t, s)$ of $K(t, s)$ has been assumed to exist, the solution to this equation can be written as

$$(2.7) \quad w'(t) = -L_t[h(t)] - \int_0^T M(t, s)L_s[h(s)]ds.$$

It again remains to be shown that $\bar{w}(0)$ and $\bar{w}(T)$ can be chosen so that the insertion of (2.7) into the right-hand side of (2.6) yields $f(t)$ identically. From the Lagrange identity (2.4) we have

$$\int_0^T r(t, s)L_s(g)ds = -g(t) + B_s(r, g),$$

where we have used $B_s(r, g)$ to denote the bilinear expression of the right-hand side of (2.4). Setting

$$g(t) = \int_0^T z(t, s)w'(s)ds,$$

(⁵) Symbolically, in obvious notation, if $h = (T_1 + T_2)m + w$, where T_{21}^{-1} exists and satisfies $T_{21}^{-1}w = 0$, $T_2T_{21}^{-1} = I + B$, it follows from $(T_1 + T_2)_{21}^{-1} = (1 + T_{21}^{-1}T_1)T_{21}^{-1}$ that $m = (1 + T_{21}^{-1}T_1)^{-1}T_{21}^{-1}h$ and that $w = h - (T_1 + T_2)m = h - [T_2(1 + T_{21}^{-1}T_1) - BT_1]m = h - T_2T_{21}^{-1}h + BT_1m = h - (I + B)h + BT_1m = B(T_1m - h)$.

we have after transposition

$$\int_0^T z(t, s) w'(s) ds = B_s \left\{ r(t, s), \int_0^T z(s, u) w'(u) du \right\} \\ - \int_0^T r(t, u) \left\{ L_u \left[\int_0^T z(u, s) w'(s) ds \right] \right\} du.$$

If the right-hand side is now substituted into (2.6), this equation becomes

$$(2.8) \quad \begin{aligned} h(t) - \bar{w}(T)r(t, T) - \bar{w}(0)r(t, 0) \\ = B_s \left\{ r(t, s), \int_0^T z(s, u) w'(u) du \right\} \\ + \int_0^T r(t, u) \left\{ w'(u) - \int_0^T L_u [z(u, v)] w'(v) dv \right\} du. \end{aligned}$$

Since $w'(t)$ is given explicitly by (2.7), the bracketed integrand of the last term of this equation can be transformed as follows:

$$(2.9) \quad \begin{aligned} w'(u) - \int_0^T L_u [z(u, s)] w'(s) ds \\ = -L_u [h(u)] - \int_0^T M(u, s) L_s [h(s)] ds \\ - \int_0^T L_u [z(u, s)] \left\{ -L_s [h(s)] - \int_0^T M(s, v) L_v [h(v)] dv \right\} ds \\ = -L_u [h(u)] - \int_0^T \left\{ M(u, s) - L_u [z(u, s)] \right. \\ \left. - \int_0^T L_u [z(u, v)] M(v, s) dv \right\} L_s [h(s)] ds. \end{aligned}$$

Since $M(t, s)$ is the resolvent kernel of $K(t, s)$, the following relation holds, where $K(t, s) = L_t [z(t, s)]$:

$$(2.10) \quad M(u, s) = K(u, s) + \int_0^T K(u, v) M(v, s) dv.$$

If this is used above, the right-hand side of (2.9) reduces to $-L_u [h(u)]$. Thus the last term of (2.8) becomes by (2.4)

$$- \int_0^T r(t, s) L_s [h(s)] ds = h(t) - B_s(r, h),$$

and (2.8) reduces to the expression

$$(2.11) \quad \bar{w}(T)r(t, T) + \bar{w}(0)r(t, 0) + B_s \left\{ r(t, s), \int_0^T z(t, u)w'(u)du - h(s) \right\} = 0.$$

This equation must be satisfied for a solution to exist where $w'(s)$ is given by (2.7). Now $w'(t)$ is a linear function of $\bar{w}(T)$ and $\bar{w}(0)$ which can be written explicitly as $w'(t) = M(t, T)\bar{w}(T) + M(t, 0)\bar{w}(0) + C(t)$, where $C(t)$ is defined as in the statement of the theorem and where (2.10) has been used again to simplify the coefficients of $\bar{w}(T)$ and $\bar{w}(0)$. Moreover, since $h(t)$ is also a linear function of $\bar{w}(T)$ and $\bar{w}(0)$ the expression

$$E(t) = \int_0^T z(t, s)w'(s)ds - h(t)$$

can be written as

$$E(t) = A(t)\bar{w}(T) + B(t)\bar{w}(0) + D(t),$$

where $A(t)$, $B(t)$, and $D(t)$ are defined in the statement of the theorem. By virtue of the fact that $r(t, s)$ is the product of two linearly independent functions, (2.11) becomes

$$\begin{aligned} \phi_2(t) \{ p(T) [\phi_1(T)E'(T) - \phi_1'(T)E(T)] + \bar{w}(T)\phi_1(T) \} \\ - \phi_1(t) \{ p(0) [\phi_2(0)E'(0) - \phi_2'(0)E(0)] + \bar{w}(0)\phi_2(0) \} = 0. \end{aligned}$$

The linear independence of the functions ϕ_1 and ϕ_2 immediately yields the linear system of the theorem, which has now been completely established.

In more explicit form the linear system can be written as

$$(2.12) \quad \begin{aligned} \bar{w}(T) + p(T) \left\{ A'(T)\bar{w}(T) + B'(T)\bar{w}(0) + D'(T) \right. \\ \left. - \frac{\phi_1'(T)}{\phi_1(T)} [A(T)\bar{w}(T) + B(T)\bar{w}(0) + D(T)] \right\} = 0, \end{aligned}$$

$$(2.13) \quad \begin{aligned} \bar{w}(0) + p(0) \left\{ A'(0)\bar{w}(T) + B'(0)\bar{w}(0) + D'(0) \right. \\ \left. - \frac{\phi_2'(0)}{\phi_2(0)} [A(0)\bar{w}(T) + B(0)\bar{w}(0) + D(0)] \right\} = 0. \end{aligned}$$

PART III. APPLICATIONS TO THE THEORY OF PREDICTION AND FILTERING

If a machine is to predict the future of a message from its past, or combine this operation with that of eliminating perturbing influences such as noise, it will be effective if and only if it is designed, not to act on one particular message, but to act on an ensemble of messages, and its effectiveness must be judged by its average performance on the messages from this en-

semble. Consequently, in the following, although actually only one message and one perturbing influence will have been observed in any given application, in theory both the message and the perturbing error term will be considered as sample functions of stochastic processes. The problem to be treated is that of finding the unbiased linear prediction of the message that has minimal variance for the entire ensemble of possible messages. Since this problem has been discussed at length in a general way in the introduction, we proceed at once to the following

THEOREM 3.1. *Let $x_a(t)$ be a continuous stochastic process with the properties:*

(a) $E[x_a(t)] = m(t)$, where $m(t) = \sum_{k=1}^n a_k g_k(t)$, where $a_k, k = 1, 2, \dots, n$, are unknown parameters and $\{g_k(t)\}, k = 1, 2, \dots, n$, are known functions;

(b) $E\{[x_a(t) - m(t)][x_a(s) - m(s)]\} = z(t, s)$;

let $e(t)$ be a continuous stochastic process such that

(c) $E[e(t)] = 0$,

(d) $E[e(t)e(s)] = r(t, s)$,

(e) $E\{e(t)[x_a(s) - m(s)]\} = 0^{(6)}$;

let a sample function of the composite process $x(t) = x_a(t) + e(t)$ be observed over an interval $0 \leq t \leq T$ and let the linear prediction of $x_a(t)$ at time $t_f > T$ be taken as

$$(3.1) \quad x_p(t_f) = \int_0^T x(t) d_t w(t, t_f).$$

Then: I. the problem of minimizing $\sigma^2 = E\{[x_p(t_f) - x(t_f)]^2\}$ subject to the unbiased condition that $E[x_p(t_f)] = \sum_{k=1}^n a_k g_k(t)$ identically in $a_k, k = 1, 2, \dots, n$, is equivalent to the free problem of minimizing the expression

$$J = \int_0^T \int_0^T [z(t, s) + r(t, s)] d_t w(t, t_f) d_s w(s, t_f) + z(t_f, t_f) - 2 \int_0^T z(t_f, t) d_t w(t, t_f) - 2 \sum_{k=1}^n \lambda_k(t_f) \left\{ \int_0^T g_k(t) d_t w(t, t_f) - g_k(t_f) \right\},$$

where $\{\lambda_k(t_f)\}, k = 1, 2, \dots, n$, are Lagrange multipliers;

II. a necessary and sufficient condition that J have a minimum is that the integral equation

$$(3.2) \quad \sum_{k=1}^n \lambda_k(t_f) g_k(t) + z(t_f, t) = \int_0^T [z(t, s) + r(t, s)] d_s w(s, t_f)$$

admit a solution for $w(t, t_f)$;

III. if problem I has a solution, the minimum value of σ^2 is given by

$$\sigma^2 = z(t_f, t_f) + \sum_{k=1}^n \lambda_k(t_f) g_k(t_f).$$

⁽⁶⁾ That is, the error process and the random part of the message are uncorrelated.

In order to prove this theorem we first recall [7] that under the hypotheses stated on $x_a(t)$ and $e(t)$ the integral (3.1) exists, and that the operations of integration and expected value may be freely interchanged. Since

$$x_p(t_f) = \int_0^T x_a(t) d_t w(t, t_f) + \int_0^T e(t) d_t w(t, t_f),$$

we have by conditions (d) and (a) that

$$\begin{aligned} m(t_f) &= E[x_p(t_f)] = \int_0^T E[x_a(t)] d_t w(t, t_f) \\ &= \int_0^T m(t) d_t w(t, t_f) = \sum_{k=1}^n a_k \int_0^T g_k(t) d_t w(t, t_f). \end{aligned}$$

Thus the requirement that the estimate be unbiased yields the isoperimetric conditions

$$(3.3) \quad \int_0^T g_k(t) d_t w(t, t_f) = g_k(t_f), \quad k = 1, 2, \dots, n.$$

Let $y(t) = x(t) - m(t)$, so that $E[y(t)y(s)] = z(t, s)$; then by a straightforward calculation

$$\begin{aligned} \sigma^2 &= E[x_p(t_f) - m(t_f) - y(t_f)]^2 \\ &= E\left\{ \int_0^T [y(t) + e(t)] d_t w(t, t_f) - y(t_f) \right\}^2 \\ &= \int_0^T \int_0^T [z(t, s) + r(t, s)] d_t w(t, t_f) d_s w(s, t_f) \\ &\quad - 2 \int_0^T z(t, t_f) d_t w(t, t_f) + z(t_f, t_f). \end{aligned}$$

An application of the usual techniques of the calculus of variations to the problem of minimizing the above expression for $w(t, t_f)$ subject to the side conditions (3.3) yields the free problem for the functional J stated in the theorem. From this free problem one obtains, in the usual way, the integral equation stated in the theorem as a necessary condition that J have a minimum. In order to see that it is also sufficient, it is convenient to introduce

$$S(t, s) = z(t, s) + r(t, s), \quad k(t, t_f) = \sum_{k=1}^n \lambda_k(t_f) g_k(t) + z(t, t_f),$$

so that the integral equation (3.2) becomes

$$k(t, t_f) = \int_0^T S(t, s) d_s w(s, t_f).$$

Let $q(t, t_f)$ denote a solution to this equation and let $w(t, t_f)$ be any other weighting function that satisfies the side conditions (3.3). Now if the terms in J involving a single integral of $w(t, t_f)$ are collected, J can be written as

$$J = \int_0^T \int_0^T S(t, s) d_t w(t, t_f) d_s w(s, t_f) + z(t_f, t_f) \\ + 2 \sum_{k=1}^n \lambda_k(t_f) g_k(t_f) - 2 \int_0^T \int_0^T S(t, s) d_s q(s, t_f) d w_t(t, t_f).$$

By completing the square, we may write

$$J = z(t_f, t_f) + 2 \sum_{k=1}^n \lambda_k(t_f) g_k(t_f) \\ + \int_0^T \int_0^T S(t, s) [d_t w(t, t_f) - d_t q(t, t_f)] [d_s w(s, t_f) - d_s q(s, t_f)] \\ - \int_0^T \int_0^T S(t, s) d_s q(s, t_f) d_t q(t, t_f).$$

Now $S(t, s)$ is a semi-definite kernel, since

$$\int_a^b \int_a^b S(t, s) f(t) f(s) dt ds = E \left\{ \int_a^b [x(t) - m(t) + e(t)] f(t) dt \right\}^2 \geq 0;$$

so J cannot have a value smaller than that given by $q(t, t_f) = w(t, t_f)$. Thus, any solution to the integral equation (3.2) yields a best value for J . If a solution to the integral equation has been obtained and the Lagrange multipliers evaluated by the side conditions (3.3), it is easy to obtain an explicit expression for the minimal variance σ^2 , which can be written as

$$\sigma^2 = \int_0^T d_t w(t, t_f) \left\{ \int_0^T S(t, s) d_s w(s, t_f) - 2z(t, t_f) \right\} + z(t_f, t_f).$$

Now if the integrand is replaced by its value as given by the integral equation (3.2), one has

$$\sigma^2 = z(t_f, t_f) + \sum_{k=1}^n \lambda_k(t_f) \int_0^T g_k(t) d_t w(t, t_f).$$

The side conditions (3.3) immediately yield the value given in Theorem 3.1.

COROLLARY 3.1. *If, in Theorem 3.1⁽⁷⁾, $x_a(t) = m(t)$, a necessary and sufficient condition that the prediction problem formulated there have a solution is that*

⁽⁷⁾ In this case the requirement that the estimate be unbiased can be replaced by the reasonable physical requirement that if $e(t) = 0$, then $x_p(t_f) = x_a(t_f)$. This constraint was employed in the auto-correlation theory and it leads to the same side conditions.

the integral equation

$$(3.4) \quad \sum_{k=1}^n \lambda_k(t_f) g_k(t) = \int_0^T r(t, s) d_s w(s, t_f)$$

have a solution for $w(t, t_f)$. In this case the minimum value of the variance is given by

$$\sigma^2 = \sum_{k=1}^n \lambda_k(t_f) g_k(t_f).$$

Corresponding results can be obtained in the event that the prediction is taken in the more general form (0.4). By way of example, we shall briefly consider the case in which $x_a(t) = m(t)$ and the number q of (0.4) is taken as unity. In particular, this type of a prediction operator implies that $e'(t)$ exists as a second-order process. In this event it is known [7] that the derivatives $\partial/\partial t$, $\partial/\partial s$, and $\partial^2/\partial t \partial s$ of $r(t, s) = E[e(t)e(s)]$ exist even for $t=s$ in the latter case, and that these derivatives are permutable with the operation of taking expected values. We therefore have the following

THEOREM 3.2. *Let $x_a(t) = m(t)$, where $m(t)$ is the same as in Theorem 3.1, and let $e(t)$ and $e'(t)$ be second-order stochastic processes such that $E[e(t)] = 0$ and $E[e(t)e(s)] = r(t, s)$; let $x(t) = x_a(t) + e(t)$; then*

I. a necessary and sufficient condition that the problem of minimizing the variance $E\{[x_p(t_f) - x(t_f)]^2\}$ subject to the condition $E[x_p(t_f)] = m(t)$ have a solution for predictors of the form

$$(3.5) \quad x_p(t_f) = \int_0^T x(t) d_t p(t, t_f) + \int_0^T x'(t) d_t q(t, t_f)$$

is that the free problem of minimizing the functional

$$\begin{aligned} F = & \int_0^T \int_0^T r(t, s) d_t p(t, t_f) d_s p(s, t_f) + \int_0^T \int_0^T \frac{\partial r(t, s)}{\partial t} d_t q(t, t_f) d_s p(s, t_f) \\ & + \int_0^T \int_0^T \frac{\partial r(t, s)}{\partial s} d_t p(t, t_f) d_s q(t, t_f) + \int_0^T \int_0^T \frac{\partial^2 r(t, s)}{\partial t \partial s} d_t q(t, t_f) d_s q(t, t_f) \\ & - 2 \sum_{k=1}^n \lambda_k(t_f) \left[\int_0^T g_k(t) d_t p(t, t_f) + \int_0^T g'_k(t) d_t q(t, t_f) - g_k(t_f) \right] \end{aligned}$$

have a solution, where $\lambda_k(t_f)$, $k=1, 2, \dots, n$, are unknown Lagrange multipliers;

II. a necessary and sufficient conditional that the functional F have a minimum is that the following system of integral equations have a solution for $p(t, t_f)$ and $q(t, t_f)$:

$$\sum_{k=1}^n \lambda_k(t_f) g_k(t) = \int_0^T r(t, s) d_s p(s, t_f) + \int_0^T \frac{\partial r(t, s)}{\partial s} d_s q(s, t_f),$$

$$\sum_{k=1}^n \lambda_k(t_f) g'_k(t) = \int_0^T \frac{\partial r(t, s)}{\partial t} d_s p(s, t_f) + \int_0^T \frac{\partial^2 r(t, s)}{\partial t \partial s} d_s q(s, t_f);$$

III. if these equations have a solution, the minimum variance is given by

$$\sigma^2 = \sum_{k=1}^n \lambda_k(t_f) g_k(t_f).$$

The proof of this theorem proceeds as that of Theorem 3.1, except that the n side conditions corresponding to (3.3) become

$$\int_0^T g_k(t) d_t p(t, t_f) + \int_0^T g'_k(t) d_t q(t, t_f) = g_k(t_f), \quad k = 1, 2, \dots, n.$$

The free problem for the functional F yields the integral equations

$$2 \sum_{k=1}^n \lambda_k(t_f) g_k(t) = 2 \int_0^T r(t, s) d_s p(s, t_f) + \int_0^T \frac{\partial r(s, t)}{\partial s} d_s q(s, t_f)$$

$$+ \int_0^T \frac{\partial r(t, s)}{\partial s} d_s q(s, t_f),$$

$$2 \sum_{k=1}^n \lambda_k(t_f) g'_k(t) = \int_0^T \frac{\partial r(t, s)}{\partial t} d_s p(s, t_f) + \int_0^T \frac{\partial r(s, t)}{\partial t} d_s p(s, t_f)$$

$$+ 2 \int_0^T \frac{\partial^2 r(t, s)}{\partial t \partial s} d_s q(s, t_f).$$

However, since $r(t, s)$ is a covariance function, it is symmetric, i.e., $r(t, s) = r(s, t)$, and from this it follows that

$$\frac{\partial r(t, s)}{\partial t} = \frac{\partial r(s, t)}{\partial t}, \quad \frac{\partial r(t, s)}{\partial s} = \frac{\partial r(s, t)}{\partial s};$$

and if these are inserted into the above equations they become the simpler and more symmetric equations stated in the theorem. In this symmetric form it is further obvious that the second equation is simply the derivative of the first with respect to t . The other details of the proof are omitted.

In order to complete the problems of prediction and filtering formulated in this part, it remains to solve the integral equations that have been derived in Theorems 3.1 and 3.2 as well as that of Corollary 3.1. If one defines

$$f_1(t, t_f) = \sum_{k=1}^n \lambda_k(t_f) g_k(t), \quad f_2(t, t_f) = \sum_{k=1}^n \lambda_k(t_f) g'_k(t) + z(t_f, t),$$

and sets up correspondences between these and the $f(t)$ of Part II as well as between $w(t, t_f)$ and $w(t)$ of that part, the integral equations (3.4) and (3.2) derived in this part can be identified with their prototypes (2.4) and (2.2) respectively of Part II. The fact that t_f is carried along as an additional parameter has no effect on the theory of that part. If the hypotheses of the theorems of that part are satisfied, as they surely are if $r(t, s)$ belongs to a stochastic process generated as in Part I, the constructive solutions obtained for the integral equations yield explicit expressions for $\partial w(t, t_f)/\partial t$ and for $\bar{w}(T, t_f)$ and $\bar{w}(0, t_f)$ in terms of the unknown Lagrange multipliers $\lambda_k(t_f)$, $k=1, 2, \dots, n$. Explicit expressions can then be found for these quantities by eliminating these unknown multipliers through the side conditions (3.3), which can be written as

$$\int_0^T g_k(t) \frac{\partial w(t, t_f)}{\partial t} dt + \bar{w}(T, t_f)g_k(T) + \bar{w}(0, t_f)g_k(0) = g_k(t_f),$$

for $k=1, 2, \dots, n$. Once this has been done, the prediction assumes the form

$$x_p(t_f) = \int_0^T x(t) \frac{\partial w(t, t_f)}{\partial t} dt + x(T)\bar{w}(T, t_f) + x(0)\bar{w}(0, t_f).$$

We shall not carry out the algebraic details in general but illustrate them in two examples at the end of this part.

It is interesting to note the results in the special case in which $x_a(t) = m(t) = 0$ and $e(t)$ is a stochastic process satisfying the hypotheses of Theorem 3.1. The problem reduces to that of the prediction of the future value of the error process itself, and the side conditions, automatically satisfied, disappear. For the prediction, one has

$$e_p(t_f) = \int_0^T e(t) d_t w(t, t_f);$$

and it is easy to see that the variance to be minimized is given by

$$\sigma^2 = r(t_f, t_f) + \int_0^T \int_0^T r(t, s) d_t w(t, t_f) d_s w(s, t_f) - 2 \int_0^T r(t_f, t) d_t w(t, t_f)$$

while the integral equation corresponding to (3.4) can be written as

$$r(t_f, t) = \int_0^T r(t, s) \frac{\partial w(s, t_f)}{\partial s} ds + r(t, T)\bar{w}(T, t_f) + r(t, 0)\bar{w}(0, t_f).$$

Using the identification process mentioned above, we find that the results of Theorem 2.1 imply $\partial w(t, t_f)/\partial t = \bar{w}(0, T) = 0$, so that in this special case the prediction reduces simply to $e_p(t_f) = e(T)\bar{w}(T, t_f)$, where $\bar{w}(T, t_f)$ is given by the expression of Theorem 2.1. This is a generalization of the well known corresponding result for the one-dimensional Markovian process. It itself

can obviously be generalized to error processes generated by equations of higher order than the first by the modified technique that has been indicated throughout.

The technique must also be modified before a solution to the system of integral equations given in Theorem 3.2 can be included in the theory developed here. We have already observed that the second of these equations is merely the derivative of the first with respect to t . The first of these equations is then replaced by the following:

$$\begin{aligned} \sum_{k=1}^n \lambda_k(t_f)g_k(t) &= \int_0^T r(t, s) \frac{\partial p(s, t_f)}{\partial s} dt + \int_0^T \frac{\partial r(t, s)}{\partial s} \frac{\partial q(s, t_f)}{\partial s} ds \\ &+ r(t, T)\bar{p}(T, t_f) + r(t, 0)\bar{p}(0, t_f) \\ &+ \frac{\partial r}{\partial \omega}(t, T)\bar{q}(T, t_f) + \frac{\partial r}{\partial \omega}(t, 0)\bar{q}(0, t_f). \end{aligned}$$

This in turn can be transformed by integration by parts of the term depending upon $q(t, t_f)$ into the form

$$\begin{aligned} \sum_{k=1}^n \lambda_k(t_f)g_k(t) &= \int_0^T r(t, s) \left[\frac{\partial p(s, t_f)}{\partial s} - \frac{\partial^2 q(s, t_f)}{\partial s^2} \right] ds \\ &+ \left[\bar{p}(T, t_f) + \frac{\partial q(T, t_f)}{\partial \alpha} \right] r(t, T) \\ &+ \left[\bar{p}(0, t_f) - \frac{\partial q(0, t_f)}{\partial \alpha} \right] r(t, 0) \\ &+ \bar{q}(T, t_f) \frac{\partial r}{\partial \omega}(t, T) + \bar{q}(0, t_f) \frac{\partial r}{\partial \omega}(t, 0). \end{aligned}$$

By defining

$$w(t, t_f) = \frac{\partial p(t, t_f)}{\partial \alpha} - \frac{\partial^2 q(t, t_f)}{\partial \alpha^2},$$

$$\bar{w}(T, t_f) = \bar{p}(T, t_f) + \frac{\partial q(T^-, t_f)}{\partial \alpha}, \quad \bar{w}(0, t_f) = \bar{p}(0, t_f) - \frac{\partial q(0^+, t_f)}{\partial \alpha},$$

we can write this as

$$\begin{aligned} \sum_{k=1}^n \lambda_k(t_f)g_k(t) &= \int_0^T r(t, s) \frac{\partial w(s, t_f)}{\partial s} ds + \bar{w}(T, t_f)r(t, T) \\ &+ \bar{w}(0, t_f)r(t, 0) + \bar{q}(T, t_f) \frac{\partial r}{\partial \omega}(t, T) + \bar{q}(0, t_f) \frac{\partial r}{\partial \omega}(t, 0), \end{aligned}$$

which can be identified with its prototype (2.5) and therefore handled as indicated in Part II. In this case, the prediction (3.5) reduces to

$$x_p(t_f) = \int_0^T x(t) \frac{\partial w(t, t_f)}{\partial t} dt + \bar{w}(T, t_f)x(T) + \bar{w}(0, t_f)x(0) \\ + \bar{q}(T, t_f)x'(T) + \bar{q}(0, t_f)x'(0).$$

We see therefore that the theory developed in this paper is capable of solving a large class of important prediction and filtering problems. We conclude with specific examples of the problems of Corollary 3.1 and of Theorem 3.1. The first of these has been treated previously by Phillips and Weiss [2] and by Cunningham and Hynd [3]. The second is a nonstationary mixed filter problem.

An example of prediction. If the errors are assumed to be a stationary stochastic process of the Markoff type, the covariance function is the Picard kernel $e^{-\beta|u-v|}$. Further, let $x_a(t)$ be specialized to a linear function of the time, say

$$x_a(t) = a_1 + a_2t.$$

Under these conditions the integral equation (3.4) specializes to

$$(3.6) \quad f(t_f, t) = \lambda_1(t_f) + \lambda_2(t_f)t = \int_0^T e^{-\beta|t-s|} \frac{\partial w(s, t_f)}{\partial s} ds \\ + \bar{w}(T, t_f)e^{-\beta(T-t)} + \bar{w}(0, t_f)e^{-\beta t},$$

while the isoperimetric side conditions (3.3) become

$$(3.7) \quad \int_0^T \frac{\partial w(s, t_f)}{\partial s} ds + \bar{w}(T, t_f) + \bar{w}(0, t_f) = 1,$$

$$(3.8) \quad \int_0^T s \frac{\partial w(s, t_f)}{\partial s} ds + T\bar{w}(T, t_f) = t_f.$$

We recall that the function $(1/2\beta)e^{-\beta|t-s|}$ is the Green's function of the differential equation

$$L_t(y) = y'' - \beta^2y = 0,$$

which satisfies

$$\lim_{t \rightarrow -\infty} y(t) = \lim_{t \rightarrow \infty} y(t) = 0.$$

To apply the general theory it is convenient to replace (3.6) by the equivalent

$$g(t, t_f) = \frac{1}{2\beta} f(t, t_f) = \frac{\lambda_1(t_f) + t\lambda_2(t_f)}{2\beta} \\ = \frac{1}{2\beta} \left[\int_0^T e^{-\beta|t-s|} \frac{\partial w(s, t_f)}{\partial s} ds + \bar{w}(T, t_f)e^{-\beta(T-t)} + \bar{w}(0, t_f)e^{-\beta t} \right].$$

Consequently by Theorem 2.1

$$\frac{\partial w(t, t_f)}{\partial t} = \frac{1}{2\beta} \left(\beta^2 g - \frac{\partial^2 g}{\partial t^2} \right) = \frac{\beta}{2} [\lambda_1(t_f) + t\lambda_2(t_f)].$$

Since

$$r(t, s) = \frac{e^{-\beta t} e^{\beta s}}{2\beta} \quad (t \geq s)$$

$$= \frac{e^{-\beta s} e^{\beta t}}{2\beta} \quad (t \leq s),$$

it follows from Theorem 2.1 that

$$(3.9) \quad \bar{w}(T, t_f) = \frac{f(T, t_f)}{2} + \frac{1}{2\beta} \frac{\partial f(T, t_f)}{\partial t} = \frac{\lambda_1(t_f)}{2} + \frac{\lambda_2(t_f)}{2} \left(T + \frac{1}{\beta} \right),$$

$$(3.10) \quad \bar{w}(0, t_f) = \frac{f(0, t_f)}{2} - \frac{1}{2\beta} \frac{\partial g(0, t_f)}{\partial t} = \frac{\lambda_1(t_f)}{2} - \frac{\lambda_2(t_f)}{2\beta}.$$

The insertion of these values into (3.7), (3.8) gives

$$(3.11) \quad \frac{\beta}{2} \int_0^T (\lambda_1 + \lambda_2 v) dv + \lambda_1 + \lambda_2 \frac{T}{2} = 1,$$

$$(3.12) \quad \frac{\beta}{2} \int_0^T (\lambda_1 v + \lambda_2 v^2) dv + \frac{\lambda_1}{2} T + \frac{\lambda_2}{2} \left(T^2 + \frac{T}{\beta} \right) = t_f.$$

Equations (3.9) to (3.12) constitute a system of four simultaneous linear equations in the four unknowns $\lambda_1(t)$, $\lambda_2(t)$, $\bar{w}(0, t_f)$, and $\bar{w}(T, t_f)$. Integrating (3.11) and (3.12) yields

$$(3.13) \quad \frac{\lambda_1}{2} (\beta T + 2) + \frac{\lambda_2}{2} \left(\frac{\beta T^2}{2} + T \right) = 1,$$

$$(3.14) \quad \frac{\lambda_1}{2} \left(\frac{\beta T^2}{2} + T \right) + \frac{\lambda_2}{2} \left(\frac{\beta T^3}{3} + \frac{T}{\beta} + T^2 \right) = t_f,$$

giving

$$(3.15) \quad \lambda_1(t_f) = \frac{8\beta T^2 + 24\beta T + 24 - 12\beta t_f(\beta T + 2)}{\beta^3 T^3 + 8\beta^2 T^2 + 24\beta T + 24},$$

$$(3.16) \quad \lambda_2(t_f) = \frac{12}{T} \frac{\beta T - 2\beta t_f}{\beta^2 T^2 + 6\beta T + 12}.$$

In order to see that these correspond to the results obtained by Cunningham and Hynd [3], one has to identify our quantities λ_1 , λ_2 , t_f , and β with their

A , B , $-T_f$, and λ , respectively. Then equations (3.13), (3.14), (3.15) check with those obtained by Cunningham and Hynd and the discrepancy in (3.16) is removed by pointing out an error in sign in [3, p. 66, line 17].

With the aid of (3.9), (3.10), one can now write the expressions for

$$\bar{w}(T, t_f) = \frac{6}{T} \cdot \frac{(5/3)\beta^2 T^3 - 5\beta T^2 + 2T - t_f(2\beta^2 T^2 + 6\beta T + 6\beta^2 T^2 + 12\beta + 4)}{\beta^3 T^3 + 8\beta^2 T^2 + 24\beta T + 24},$$

$$\bar{w}(0, t_f) = \frac{6}{T} \cdot \frac{(2/3)\beta^2 T^3 - \beta T^2 + 2\beta T - 2T + 2 - t_f(\beta^2 T - 2\beta T + 2\beta - 4)}{\beta^3 T^3 + 8\beta^2 T^2 + 24\beta T + 24};$$

for the derivative of the weighting function from (7.5),

$$\frac{\partial w(t, t_f)}{\partial t} = \frac{\beta}{2} (\lambda_1 + t\lambda_2)$$

$$= \frac{\beta}{2} \cdot \frac{8\beta^2 T^2 + 24\beta T + 24 - 12\beta t_f(\beta T + 2)(1 + 2t/T - t/t_f)}{\beta^3 T^3 + 8\beta^2 T^2 + 24\beta T + 24};$$

and for the predicted variance from Theorem 2.1,

$$\sigma_x^2(t) = \lambda_1(t) + t_f \lambda_2(t) = \frac{8\beta^2 T^2 + 24\beta T + 24 - (24\beta t_f^2/T)(\beta T + 2)}{\beta^3 T^3 + 8\beta^2 T^2 + 24\beta T + 24}.$$

It is worth noting that as $\beta \rightarrow \infty$, i.e., as $r(t, s) \rightarrow 0$, the parameters λ_1, λ_2 vanish as

$$\lambda_1(t_f) \rightarrow \frac{8\beta^2 T^2 - 12\beta^2 T t_f}{\beta^3 T^3} \rightarrow 0,$$

$$\lambda_2(t_f) \rightarrow \frac{12}{T} \frac{\beta T - 2\beta t_f}{\beta^2 T^2} \rightarrow 0,$$

and as $\bar{w}(0, t_f), \bar{w}(T, t_f) \rightarrow 0$.

The "asymptotic" weighting function, however, becomes

$$\lim_{t \rightarrow \infty} \frac{\partial w(t, t_f)}{\partial t} = \frac{2}{T^3} [2T^2 - 3tT + 3t_f(2t - T)],$$

in such a manner that the isoperimetric side conditions (3.7) and (3.8) hold.

In this example it is not difficult to verify that the same results would have been obtained if one had attempted to estimate a_1, a_2 by assuming estimators $A_i, i = 1, 2$, of the form

$$A_i = \int_0^T x(t) d_i p_i(t, t_f),$$

and requiring that these estimates be unbiased. This imposes the following

conditions on $p_i(t, t_f)$ for $i=1, 2$:

$$\int_0^T d_t p_i(t, t_f) = \delta_{i1}, \quad \int_0^T t d_t p_i(t, t_f) = t_f \delta_{i2},$$

where we have employed the Kronecker delta. The prediction would then be the deterministic expression $x_p(t_f) = A_1 + A_2 t_f$, so that this problem can be interpreted as a "smoothing problem" in which A_1 and A_2 are the results of the smoothing operation applied to the interval $0 \leq t \leq T$. This interpretation is easier to mechanize by simulation in the stationary case. One can of course obtain this same interpretation for any of the problems formulated in Corollary 3.1 or Theorem 3.2 and their generalizations.

A nonstationary filter problem. Let it be assumed that the message consists of Brownian random fluctuations about a mean linear path and that, as in the previous example, the error process is stationary and Markovian. We again assume that the message and error processes are uncorrelated. Let⁽⁸⁾

$$\begin{aligned} r(t, s) &= E[e(t)e(s)] = e^{-\beta|t-s|}, \\ E[x_a(t)] &= m(t) = a_1 + a_2 t, \\ z(t, s) &= \sigma^2 \min \{ |t|, |s| \}; \end{aligned}$$

as before, let it be assumed that the sum of a sample message function and a sample error function has been observed throughout the interval $0 \leq t \leq T$. In terms of the notation of Part II, the integral equation corresponding to (2.6) becomes

$$\begin{aligned} (3.17) \quad h(t, t_f) &= \lambda_1(t_f) + \lambda_2(t_f)t + z(t_f, t) - \bar{w}(T, t_f)\sigma^2 t \\ &= \int_0^T [z(t, s) + e^{-\beta|t-s|}] \frac{\partial w(s, t_f)}{\partial s} ds + \bar{w}(T, t_f)e^{-\beta|T-t|} \\ &\quad + \bar{w}(0, t_f)e^{-\beta t}. \end{aligned}$$

Now let

$$L_t = \frac{D^2}{2\beta + \sigma^2} - \gamma^2, \quad \gamma^2 = \frac{\beta^2}{2\beta + \sigma^2}, \quad D = \frac{\partial}{\partial t},$$

and

$$-L_t \{ h(t, t_f) \} = f(t, t_f).$$

By the application of the operator L_t to equation (3.17) for $0 < t < T$,

$$(3.18) \quad \frac{\partial w(t, t_f)}{\partial t} = f(t, t_f) - \gamma^2 \int_0^T z(t, s) \frac{\partial w(s, t_f)}{\partial s} ds.$$

⁽⁸⁾ With the usual normalization $x(0) = 0$ and $\sigma^2 = 1$, $z(t, s)$ is the covariance of a ξ -process.

In order to solve this Fredholm integral equation of the second kind, we observe that any solution to it must satisfy the differential equation

$$(D_t^2 - \gamma^2 \sigma^2) \frac{\partial w(t, t_f)}{\partial t} = D_t^2 f(t, t_f),$$

as follows directly below:

$$\begin{aligned} \frac{\partial w(t, t_f)}{\partial t} &= f(t, t_f) - \gamma^2 \sigma^2 \int_0^T v m(v, t_f) dv - \gamma^2 \sigma^2 t \int_0^T m(v, t_f) dv, \\ \frac{\partial^2 w(t, t_f)}{\partial t^2} &= \frac{\partial f(t, t_f)}{\partial t} - \gamma^2 \sigma^2 \int_t^T m(v, t_f) dv, \\ \frac{\partial^3 w(t, t_f)}{\partial t^3} &= \frac{\partial^2 f(t, t_f)}{\partial t^2} + \gamma^2 \sigma^2 \frac{\partial w(t, t_f)}{\partial t}. \end{aligned}$$

Moreover, these equations imply that

$$\frac{\partial w(0, t_f)}{\partial t} = f(0, t_f), \quad \frac{\partial^2 w(T, t_f)}{\partial t^2} = \frac{\partial f(T, t_f)}{\partial t}.$$

Now the solution to (3.18) subject to these last two conditions may be found to be

$$(3.19) \quad \frac{\partial w(t, t_f)}{\partial t} = f(t, t_f) + \int_0^T M(t, s) f(s, t_f) ds,$$

where $\alpha^2 = \gamma^2 \sigma^2$, and

$$(3.20) \quad M(t, s) = \frac{\alpha}{2} \frac{1}{\cosh(\alpha t)} \left\{ \sinh[\alpha(T - t - s)] + e^{-\alpha T} \cosh[\alpha(t - s)] - \cosh \alpha T e^{-\alpha|t-s|} \right\}.$$

This is in the form (2.7) and it is readily verified that the kernel

$$z(t, s) + e^{-\beta|t-s|}$$

satisfies the hypothesis of the usual Fredholm theory; it follows that (3.19) and (3.20) give the unique solution to (3.18). It should be observed that $M(t, s)$ is symmetric in t, s .

To use $M(t, s)$ as an integrand, it is necessary to distinguish the two cases

$$\begin{aligned} M_1(t, s) &= - \frac{\alpha \sinh(\alpha t) \cosh[\alpha(T - s)]}{\cosh(\alpha T)}, & s > t, \\ M_2(t, s) &= - \frac{\alpha \sinh(\alpha s) \cosh[\alpha(T - t)]}{\cosh(\alpha T)}, & s < t. \end{aligned}$$

For the case of prediction, when $t_f > T$, the formulas of Theorem 3.1 are as follows:

$$M(t, T) = - \frac{\alpha \sinh (\alpha t)}{\cosh (\alpha T)},$$

$$M(t, 0) = 0,$$

$$C(t, t_f) = \gamma^2 \lambda_1 \frac{\cosh [\alpha(T-t)]}{\cosh (\alpha T)} + \gamma^2 (\lambda_2 + \sigma^2) \frac{\sinh (\alpha t)}{\alpha \cosh (\alpha T)},$$

$$A(t, t_f) = \frac{\sigma^2}{2} \frac{\sinh (\alpha t)}{\cosh (\alpha T)},$$

$$A(0, t_f) = 0,$$

$$\frac{\partial A(0, t_f)}{\partial t} = \frac{\sigma^2}{\cosh (\alpha T)},$$

$$A(T, t_f) = \frac{\sigma^2}{\alpha} \tanh (\alpha T),$$

$$\frac{\partial A(T, t_f)}{\partial t} = \sigma^2,$$

$$B(t, t_f) = B(0, t_f) = B(T, t_f) = \frac{\partial B(0, t_f)}{\partial t} = \frac{\partial B(T, t_f)}{\partial t} = 0,$$

$$D(0, t_f) = F_0 = -\lambda_1,$$

$$D(T, t_f) = -\frac{\lambda_2 + \sigma^2}{\alpha} \tanh (\alpha T) - \frac{\lambda_1}{\cosh (\alpha T)},$$

$$\frac{\partial D(0, t_f)}{\partial t} = \frac{1}{\cosh (\alpha T)} [\alpha \sinh (\alpha T) - \lambda_2 - \sigma^2],$$

$$\frac{\partial D(T, t_f)}{\partial t} = -\lambda_2 - \sigma^2.$$

From the equation

$$\frac{\partial w(t, t_f)}{\partial t} = M(t, T) \bar{w}(T, t_f) + M(t, 0) \bar{w}(0, t_f) + C(t, t_f)$$

we have

$$\frac{\partial w(t, t_f)}{\partial t} = -\frac{\alpha \sinh (\alpha t)}{\cosh (\alpha T)} w(T, t_f) + C(t, t_f).$$

The conditions (2.12), (2.13) and the isoperimetric side conditions

$$\int_0^T d_t w(t, t_f) = \int_0^T \frac{\partial w(t, t_f)}{\partial t} dt + \bar{w}(0, t_f) + \bar{w}(T, t_f) = 1,$$

$$\int_0^T t d_t w(t, t_f) = \int_0^T t \frac{\partial w(t, t_f)}{\partial t} dt + T \bar{w}(T, t_f) = t_f$$

reduce the problem to the solution of the following four linear equations:

$$(3.21) \quad \alpha \sinh(\alpha T) \cdot \lambda_1 + [\cosh(\alpha T) - 1] \lambda_2 + \sigma^2 \bar{w}(T, t_f) + \sigma^2 \cosh(\alpha T) \bar{w}(0, t_f) = \sigma^2,$$

$$(3.22) \quad [\cosh(\alpha T) - 1] \lambda_1 + \left[T \cosh(\alpha T) - \frac{\sinh(\alpha T)}{\alpha} \right] \lambda_2 + \frac{\sigma^2}{\alpha} \sinh(\alpha T) \cdot \bar{w}(T, t_f) = \frac{\sigma^2}{\alpha} \sinh(\alpha T) + \sigma^2(t_f - T) \cosh(\alpha T),$$

$$(3.23) \quad -\beta \lambda_1 - \left[\frac{\beta \sinh(\alpha T)}{\alpha} + \cosh(\alpha T) \right] \lambda_2 + \left[(2\beta + 2\sigma^2) \cosh(\alpha T) + \frac{\beta \sigma^2}{\alpha} \sinh(\alpha T) \right] \bar{w}(T, t_f) = \sigma^2 \cosh(\alpha T) + \frac{\beta \sigma^2}{\alpha} \sinh(\alpha T),$$

$$(3.24) \quad [\beta \cosh(\alpha T) + \alpha \sinh(\alpha T)] \lambda_1 - \lambda_2 + \sigma^2 \bar{w}(T, t_f) - (2\beta + \sigma^2) \cosh(\alpha T) \bar{w}(T, t_f) = \sigma^2.$$

For the case of pure filtering, i.e., for $t_f < T$, the four corresponding equations are given here without proof. The difference is due to the discontinuity of $M(t, s)$ at $t = s$. The equations are:

$$\alpha \sinh(\alpha T) \lambda_1 + [\cosh(\alpha T) - 1] \lambda_2 + \sigma^2 \bar{w}(T, t_f) + \sigma^2 \cosh(\alpha T) \bar{w}(0, t_f) = \sigma^2 \cosh[\alpha(T - t_f)],$$

$$[\cosh(\alpha T) - 1] \lambda_1 + \left[T \cosh(\alpha T) - \frac{\sinh(\alpha T)}{\alpha} \right] \lambda_2 + \frac{\sigma^2}{\alpha} \sinh(\alpha T) \bar{w}(T, t_f) = \frac{\sigma^2}{\alpha} \sinh(\alpha t_f),$$

$$-\beta \lambda_1 - \left[\frac{\beta}{\alpha} \sinh(\alpha T) + \cosh(\alpha T) \right] \lambda_2 + \left[(2\beta + 2\sigma^2) \cosh(\alpha T) + \frac{\beta \sigma^2}{\alpha} \sinh(\alpha T) \right] \bar{w}(T, t_f) = \frac{\beta \sigma^2}{\alpha} \sinh(\alpha t_f),$$

$$[\beta \cosh(\alpha T) + \alpha \sinh(\alpha T)] \lambda_1 - \lambda_2 + \sigma^2 \bar{w}(T, t_f) - (2\beta + \sigma^2) \cosh(\alpha T) \bar{w}(0, t_f) = \sigma^2 \cosh[\alpha(T - t_f)].$$

These can be solved explicitly for the unknowns of the problem, $\lambda_1(t_f)$, $\lambda_2(t_f)$, $\bar{w}(0, t_f)$, and $\bar{w}(T, t_f)$. The solution is too lengthy to be given here explicitly, but can be obtained easily for special cases.

A partial check on the correctness of equations (3.21) to (3.24) is obtained by letting $\sigma^2 \rightarrow 0$, so that $\alpha \rightarrow 0$. In this case the equations reduce to a set of simultaneous equations, identical to (3.9) to (3.12) obtained for the preceding example.

Finally by Theorem 3.1, one has the variance for $t_f > T$:

$$\sigma_x^2(t_f) = t_f[\sigma^2 + \lambda_2(t_f)] + \lambda_1(t_f).$$

REFERENCES

1. N. Wiener, *The extrapolation, interpolation and smoothing of stationary time series*, Wiley, 1946.
2. A. Kolmogoroff, *Interpolation and extrapolation*, Bull. Acad. Sci. URSS. Sér Math. vol. 5 (1941) pp. 3-14.
3. R. S. Phillips and P. Weiss, *Theoretical calculation on best smoothing of position data for gunnery prediction*, NDRC Report 532, Massachusetts Institute of Technology Radiation Laboratory, 1944.
4. L. Cunningham and W. Hynd, *Random processes in problems of air warfare*, Supplement No. 1 to the Journal of the Royal Statistical Society (1946) pp. 62-85.
5. J. L. Doob, *The elementary Gaussian processes*, Ann. of Math. Statist. (1944) pp. 282-292.
6. R. Ince, *Ordinary differential equations*, New York, Dover, 1944, Chapter XI.
7. P. Loève, *Fonctions aléatoires du second ordre*, A note in P. Lévy, *Processus stochastiques et mouvement Brownien*, Paris, Gauthier-Villars, 1948.
8. G. E. Uhlenbeck and H. T. Wang, *On the theory of Brownian motion*, Review of Modern Physics vol. 17, nos. 22-23, pp. 323-342.
9. R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, New York, Interscience, 1943, Vol. I, Chapter V.
10. R. P. Agnew, *Differential equations*, New York, McGraw-Hill, 1942, p. 174.
11. H. W. Bode and C. E. Shannon, *A simplified derivation of linear least square smoothing and prediction theory*, Proceedings of the Institute of Radio Engineers vol. 38 (1950) pp. 417-425.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.
PRINCETON UNIVERSITY,
PRINCETON, N. J.