From Papulis

11-3. The Wiener-Kolmogoroff theory

In the examples of the preceding section, the available data consisted of one, two, or at most a countable number of random variables. We now come to our main objective, namely, the estimation problem, where the data are available over an entire interval. We are given two processes $\mathbf{g}(t)$ and $\mathbf{x}(t)$. The first process is the signal $\mathbf{s}(t)$ or a functional of this signal. The second process $\mathbf{x}(t)$ is statistically related to the first, and we assume that it is "known" (see footnote, page 387) for every t in an interval (a,b), where the end points of this interval might depend on t.

We want to estimate $\mathbf{g}(t)$ for a *specific* t by a linear combination of the known values

$$\mathbf{x}(\xi) \qquad a \le \xi \le b$$

of $\mathbf{x}(t)$. This means that we seek suitable weights $h(\xi)$ such that with

$$\mathbf{g}(t) \sim \int_a^b h(\xi)\mathbf{x}(\xi)\, d\xi = \lim_{\Delta\xi \to 0} \sum_{i=1}^{n} h(\xi_i)\mathbf{x}(\xi_i)\,\Delta\xi \tag{11-27}$$

the error

$$e = E\{[\mathbf{g}(t) - \int_a^b h(\xi)\mathbf{x}(\xi)\, d\xi]^2\} \tag{11-28}$$

is minimum.

The integral in (11-27) is the limit of a sum; thus (11-27) can be viewed as an estimation of $\mathbf{g}(t)$ by a linear combination of the r.v. $\mathbf{x}(\xi_i)$ multiplied by the constants $h(\xi_i)\,\Delta\xi$. This is identical with the problem of Sec. 11-2, where now the constants a_i are $h(\xi_i)\,\Delta\xi$. From the orthogonality principle follows that these constants must be so chosen that the difference between $\mathbf{g}(t)$ and the estimation sum (integral) is orthogonal to the data; i.e.,

$$E\{[\mathbf{g}(t) - \int_a^b \mathbf{x}(\alpha)h(\alpha)\, d\alpha]\mathbf{x}(\xi)\} = 0 \qquad a \le \xi \le b \tag{11-29}$$

for every ξ in the interval (a,b). With

$$E\{\mathbf{g}(t)\mathbf{x}(\xi)\} = R_{gx}(t-\xi) \qquad E\{\mathbf{x}(\alpha)\mathbf{x}(\xi)\} = R_{xx}(\alpha - \xi)$$

the above gives our *final result*

$$R_{gx}(t-\xi) = \int_a^b R_{xx}(\alpha - \xi)h(\alpha)\, d\alpha \qquad a \le \xi \le b \tag{11-30}$$

Thus the weights $h(\xi)$ must be so selected as to satisfy the above equation. The resulting minimum m.s. error equals the expected value of the product of the error times the quantity $\mathbf{g}(t)$ to be estimated [see (11-17)]:

$$e = E\{[\mathbf{g}(t) - \int_a^b \mathbf{x}(\alpha)h(\alpha)\, d\alpha]\mathbf{g}(t)\} = R_{gg}(0) - \int_a^b R_{gx}(t-\alpha)h(\alpha)\, d\alpha \tag{11-31}$$

This is the essence of the linear m.s. estimation problem.† What remains is to identify the processes $\mathbf{g}(t)$ and $\mathbf{x}(t)$ and the interval (a,b) in a particular problem and to solve the integral equation (11-30) for the unknown function $h(\xi)$. The analytical difficulties lie in the solution of this equation.

11-4. The filtering problem

We are given the processes $\mathbf{s}(t)$ (signal) and

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{n}(t)$$

(signal plus noise). We assume that the data $\mathbf{x}(t)$ are "known" for every t from $-\infty$ to ∞. We want to estimate $\mathbf{s}(t)$ by a linear operation on these data. This problem is a special case of (11-27), with

$$\mathbf{g}(t) = \mathbf{s}(t) \qquad a = -\infty \qquad b = \infty$$

$$\mathbf{s}(t) \sim \int_{-\infty}^{\infty} h(t;\xi)\mathbf{x}(\xi)\, d\xi$$

From the stationarity of the given processes $\mathbf{s}(t)$ and $\mathbf{x}(t)$ follows that $h(t;\xi)$ should depend only on the difference $t-\xi$ (see Example 11-9):

$$\mathbf{s}(t) \sim \int_{-\infty}^{\infty} h(t-\xi)\mathbf{x}(\xi)\, d\xi = \int_{-\infty}^{\infty} \mathbf{x}(t-\alpha)h(\alpha)\, d\alpha$$

Thus our estimator

$$\hat{\mathbf{s}}(t) = \int_{-\infty}^{\infty} \mathbf{x}(t-\alpha)h(\alpha)\, d\alpha$$

can be viewed as the output of a linear time-invariant system with input $\mathbf{x}(t)$ and impulse response the unknown weights $h(t)$ (Fig. 11-2). We

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{G}(\tau)\mathbf{u}(\tau)d\tau \tag{3}$$

where we call $\mathbf{\Phi}(t, t_0)$ the *transition matrix* of (1). The transition matrix is a nonsingular matrix satisfying the differential equation

$$d\mathbf{\Phi}/dt = \mathbf{F}(t)\mathbf{\Phi} \tag{4}$$

(any such matrix is a *fundamental matrix* [23, Chapter 3]), made *unique* by the additional requirement that, for all t_0,

$$\mathbf{\Phi}(t_0, t_0) = \mathbf{I} = \text{unit matrix} \tag{5}$$

The following properties are immediate by the existence and uniqueness of solutions of (1):

$$\mathbf{\Phi}^{-1}(t_1, t_0) = \mathbf{\Phi}(t_0, t_1) \quad \text{for all} \quad t_0, t_1 \tag{6}$$

$$\mathbf{\Phi}(t_2, t_0) = \mathbf{\Phi}(t_2, t_1)\mathbf{\Phi}(t_1, t_0) \quad \text{for all} \quad t_0, t_1, t_2 \tag{7}$$

If $\mathbf{F}$ = const, then the transition matrix can be represented by the well-known formula

$$\mathbf{\Phi}(t, t_0) = \exp \mathbf{F}(t - t_0) = \sum_{i=0}^{\infty} [\mathbf{F}(t - t_0)]^i/i! \tag{8}$$

which is quite convenient for numerical computations. In this special case, one can also express $\mathbf{\Phi}$ analytically in terms of the eigenvalues of $\mathbf{F}$, using either linear algebra [22] or standard transfer-function techniques [14].

In some cases, it is convenient to replace the right-hand side of (3) by a notation that focuses attention on how the state of the system "moves" in the state space as a function of time. Thus we write the left-hand side of (3) as

$$\mathbf{x}(t) \equiv \boldsymbol{\phi}(t;\ \mathbf{x}, t_0;\ \mathbf{u}) \tag{9}$$

Read: The state of the system (1) at time t, evolving from the initial state $\mathbf{x} = \mathbf{x}(t_0)$ at time t_0 under the action of a *fixed* forcing function $\mathbf{u}(t)$. For simplicity, we refer to $\boldsymbol{\phi}$ as the *motion* of the dynamical system

4 Statement of Problem

We shall be concerned with the continuous-time analog of Problem I of reference [11], which should be consulted for the physical motivation of the assumptions stated below.

(A_1) The *message* is a random process $\mathbf{x}(t)$ generated by the *model*

$$d\mathbf{x}/dt = \mathbf{F}(t)\mathbf{x} + \mathbf{G}(t)\mathbf{u}(t) \tag{10}$$

The *observed signal* is

$$\mathbf{z}(t) = \mathbf{y}(t) + \mathbf{v}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{v}(t) \tag{11}$$

The functions $\mathbf{u}(t)$, $\mathbf{v}(t)$ in (10–11) are independent random processes (white noise) with identically zero means and covariance matrices

$$\begin{aligned} \text{cov}\,[\mathbf{u}(t), \mathbf{u}(\tau)] &= \mathbf{Q}(t)\cdot\delta(t - \tau) \\ \text{cov}\,[\mathbf{v}(t), \mathbf{v}(\tau)] &= \mathbf{R}(t)\cdot\delta(t - \tau) \quad \text{for all} \quad t, \tau \\ \text{cov}\,[\mathbf{u}(t), \mathbf{v}(\tau)] &= \mathbf{0} \end{aligned} \tag{12}$$

where δ is the Dirac delta function, and $\mathbf{Q}(t)$, $\mathbf{R}(t)$ are symmetric, nonnegative definite matrices continuously differentiable in t.

We introduce already here a restrictive assumption, which is needed for the ensuing theoretical developments:

(A_2) The matrix $\mathbf{R}(t)$ is positive definite for all t. Physically, this means that no component of the signal can be measured exactly.

To determine the random process $\mathbf{x}(t)$ uniquely, it is necessary to add a further assumption. This may be done in two different ways:

(A_3) The dynamical system (10) has reached "steady-state" under the action of $\mathbf{u}(t)$, in other words, $\mathbf{x}(t)$ is the random function defined by

$$\mathbf{x}(t) = \int_{-\infty}^{t} \mathbf{\Phi}(t, \tau)\mathbf{G}(\tau)\mathbf{u}(\tau)d\tau \tag{13}$$

This formula is valid if the system (10) is uniformly asymptotically stable (for precise definition, valid also in the nonconstant case, see [21]). If, in addition, it is true that $\mathbf{F}$, $\mathbf{G}$, $\mathbf{Q}$ are constant, then $\mathbf{x}(t)$ is a stationary random process—this is one of the chief assumptions of the original Wiener theory.

However, the requirement of asymptotic stability is inconvenient in some cases. For instance, it is not satisfied in Example 5, which is a useful model in some missile guidance problems. Moreover, the representation of random functions as generated by a linear dynamical system is already an appreciable restriction and one should try to avoid making any further assumptions. Hence we prefer to use:

(A_3') The measurement of $\mathbf{z}(t)$ starts at some fixed instant t_0 of time (which may be $-\infty$), at which time cov$[\mathbf{x}(t_0), \mathbf{x}(t_0)]$ is known.

Assumption (A_3) is obviously a special case of (A_3'). Moreover, since (10) is not necessarily stable, this way of proceeding makes it possible to treat also situations where the message variance grows indefinitely, which is excluded in the conventional theory.

The main object of the paper is to study the

OPTIMAL ESTIMATION PROBLEM. *Given known values of $\mathbf{z}(\tau)$ in the time-interval $t_0 \leqq \tau \leqq t$, find an estimate $\hat{\mathbf{x}}(t_1|t)$ of $\mathbf{x}(t_1)$ of the form*

$$\hat{\mathbf{x}}(t_1|t) = \int_{t_0}^{t} \mathbf{A}(t_1, \tau)\mathbf{z}(\tau)d\tau \tag{14}$$

(where $\mathbf{A}$ is an $n \times p$ matrix whose elements are continuously differentiable in both arguments) with the property that the expected squared error in estimating any linear function of the message is minimized:

$$\mathcal{E}[\mathbf{x}^*, \mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1|t)]^2 = \text{minimum for all } \mathbf{x}^* \tag{15}$$

Remarks. (*a*) Obviously this problem includes as a special case the more common one in which it is desired to minimize

$$\mathcal{E}\|\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1|t)\|^2$$

(*b*) In view of (A_1), it is clear that $\mathcal{E}\mathbf{x}(t_1) = \mathcal{E}\hat{\mathbf{x}}(t_1|t) = \mathbf{0}$. Hence $[\mathbf{x}^*, \hat{\mathbf{x}}(t_1|t)]$ is the *minimum variance linear unbiased estimate* of the value of any *costate* $\mathbf{x}^*$ at $\mathbf{x}(t_1)$.

(*c*) If $\mathcal{E}\mathbf{u}(t)$ is unknown, we have a more difficult problem which will be considered in a future paper.

(*d*) It may be recalled (see, e.g., [11]) that if $\mathbf{u}$ and $\mathbf{v}$ are gaussian, then so are also $\mathbf{x}$ and $\mathbf{z}$, and therefore the best estimate will be of the type (14). Moreover, the same estimate will be best not only for the loss function (15) but also for a wide variety of other loss functions.

(*e*) The representation of white noise in the form (12) is not rigorous, because of the use of delta "functions." But since the delta function occurs only in integrals, the difficulty is easily removed as we shall show in a future paper addressed to mathematicians. All other mathematical developments given in the paper are rigorous.

The solution of the estimation problem under assumptions (A_1), (A_2), (A_3') is stated in Section 7 and proved in Section 8.

5 The Dual Problem

It will be useful to consider now the dual of the optimal estimation problem which turns out to be the optimal regulator problem in the theory of control.

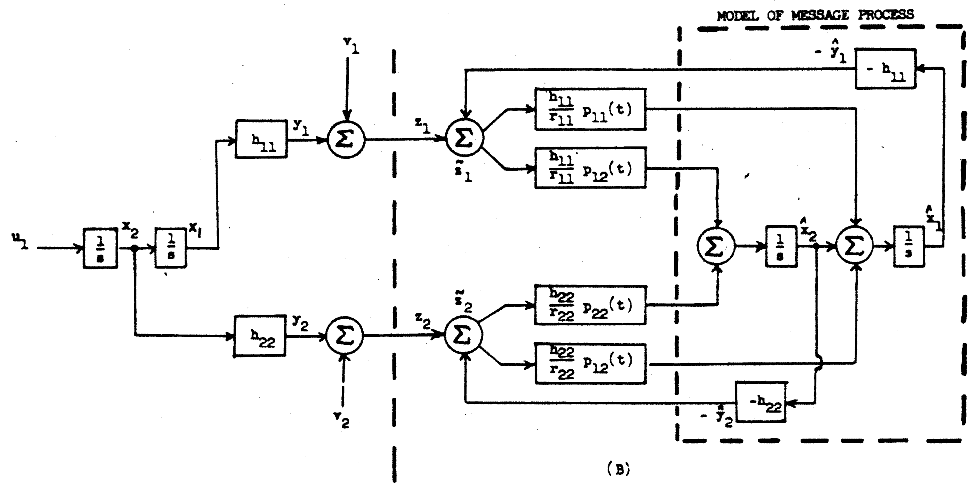

Fig. 8 Example 6: Block diagram of message process and optimal filter

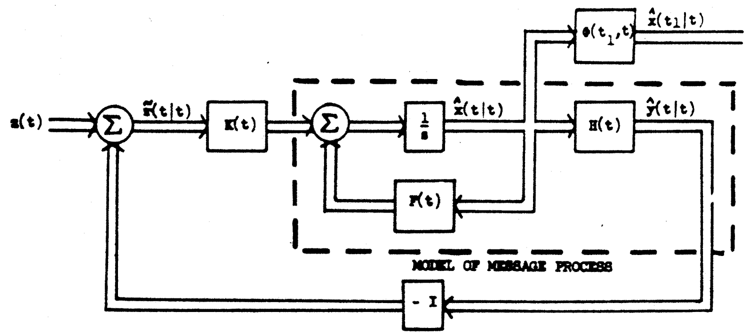

Fig. 9 General block diagram of optimal filter

$$\mathbf{G} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} h_{11} & 0 \\ 0 & h_{22} \end{bmatrix}$$

This problem was studied by Hanson [9] and Bucy [25, 26]. The dual problem is very similar to Examples 3 and 4.

7 Summary of Results: Mathematics

Here we present the main results of the paper in precise mathematical terms. At the present stage of our understanding of the problem, the rigorous proof of these facts is quite complicated, requiring advanced and unconventional methods; they are to be found in Sections 8–10. After reading this section, one may pass without loss of continuity to Section 11 which contains the solutions of the examples.

(1) *Canonical form of the optimal filter.* The optimal estimate $\hat{\mathbf{x}}(t|t)$ is generated by a linear dynamical system of the form

$$\begin{aligned} d\hat{\mathbf{x}}(t|t)/dt &= \mathbf{F}(t)\hat{\mathbf{x}}(t|t) + \mathbf{K}(t)\tilde{\mathbf{z}}(t|t) \\ \tilde{\mathbf{z}}(t|t) &= \mathbf{z}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t|t) \end{aligned} \qquad \text{(I)}$$

The initial state $\hat{\mathbf{x}}(t_0|t_0)$ of (I) is zero.

For optimal extrapolation, we add the relation

$$\hat{\mathbf{x}}(t_1|t) = \mathbf{\Phi}(t_1, t)\hat{\mathbf{x}}(t|t) \qquad (t_1 \geq t) \qquad \text{(V)}$$

No similarly simple formula is known at present for interpolation ($t_1 < t$).

The block diagram of (I) and (V) is shown in Fig. 9. The variables appearing in this diagram are vectors and the "boxes" represent matrices operating on vectors. Otherwise (except for the noncommutativity of matrix multiplication) such generalized block diagrams are subject to the same rules as ordinary block diagrams. The fat lines indicating direction of signal flow serve as a reminder that we are dealing with multiple rather than single signals.

The optimal filter (I) is a feedback system. It is obtained by taking a copy of the model of the message process (omitting the constraint at the input), forming the error signal $\tilde{\mathbf{z}}(t|t)$ and feeding the error forward with a gain $\mathbf{K}(t)$. Thus the specification of the optimal filter is equivalent to the computation of the optimal time-varying gains $\mathbf{K}(t)$. This result is general and does not depend on constancy of the model.

(2) *Canonical form for the dynamical system governing the optimal error.* Let

$$\tilde{\mathbf{x}}(t|t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t|t) \qquad (22)$$

Except for the way in which the excitations enter the optimal error, $\tilde{\mathbf{x}}(t|t)$ is governed by the same dynamical system as $\hat{\mathbf{x}}(t|t)$:

$$d\tilde{\mathbf{x}}(t|t)/dt = \mathbf{F}(t)\tilde{\mathbf{x}}(t|t) + \mathbf{G}(t)\mathbf{u}(t) - \mathbf{K}(t)[\mathbf{v}(t) + \mathbf{H}(t)\tilde{\mathbf{x}}(t|t)] \qquad \text{(II)}$$

See Fig. 10.

(3) *Optimal gain.* Let us introduce the abbreviation:

$$\mathbf{P}(t) = \operatorname{cov}[\tilde{\mathbf{x}}(t|t), \tilde{\mathbf{x}}(t|t)] \qquad (23)$$

Then it can be shown that

$$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}'(t)\mathbf{R}^{-1}(t) \qquad \text{(III)}$$

(4) *Variance equation.* The only remaining unknown is $\mathbf{P}(t)$. It can be shown that $\mathbf{P}(t)$ must be a solution of the matrix differential equation

$$d\mathbf{P}/dt = \mathbf{F}(t)\mathbf{P} + \mathbf{P}\mathbf{F}'(t) - \mathbf{P}\mathbf{H}'(t)\mathbf{R}^{-1}(t)\mathbf{H}(t)\mathbf{P} + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}'(t) \qquad \text{(IV)}$$

This is the *variance equation;* it is a system of $n(n+1)/2$[4] nonlinear differential equations of the first order, and is of the *Riccati* type well known in the calculus of variations [17, 18].

(5) *Existence of solutions of the variance equation.* Given any fixed initial time t_0 and a nonnegative definite matrix $\mathbf{P}_0$, (IV) has a unique solution

$$\mathbf{P}(t) = \mathbf{\Pi}(t;\ \mathbf{P}_0, t_0) \tag{24}$$

defined for all $|t - t_0|$ *sufficiently small,* which takes on the value $\mathbf{P}(t_0) = \mathbf{P}_0$ at $t = t_0$. This follows at once from the fact that (IV) satisfies a Lipschitz condition [21].

Since (IV) is nonlinear, we cannot of course conclude without further investigation that a solution $\mathbf{P}(t)$ exists for *all* t [21]. By taking into account the problem from which (IV) was derived, however, it can be shown that $\mathbf{P}(t)$ in (24) is defined for all $t \geqq t_0$.

These results can be summarized by the following theorem, which is the analogue of Theorem 3 of [11] and is proved in Section 8:

THEOREM 1. *Under Assumptions* (A_1), (A_2), (A_3'), *the solution of the optimal estimation problem with* $t_0 > -\infty$ *is given by relations* (I–V). *The solution* $\mathbf{P}(t)$ *of* (IV) *is uniquely determined for all* $t \geqq t_0$ *by the specification of*

$$\mathbf{P}_0 = \operatorname{cov}[\mathbf{x}(t_0), \mathbf{x}(t_0)];$$

knowledge of $\mathbf{P}(t)$ *in turn determines the optimal gain* $\mathbf{K}(t)$. *The initial state of the optimal filter is* $\mathbf{0}$.

(6) *Variance of the estimate of a costate.* From (23) we have immediately the following formula for (15):

$$\mathcal{E}[\mathbf{x}^*, \tilde{\mathbf{x}}(t|t)]^2 = \|\mathbf{x}^*\|^2_{\mathbf{P}(t)} \tag{25}$$

(7) *Analytic solution of the variance equation.* Because of the close relationship between the Riccati equation and the calculus of variations, a closed-form solution of sorts is available for (IV). The easiest way of obtaining it is as follows [17]:

Introduce the quadratic *Hamiltonian* function

$$\mathcal{H}(\mathbf{x}, \mathbf{w}, t) = -(1/2)\|\mathbf{G}'(t)\mathbf{x}\|^2_{\mathbf{Q}(t)} - \mathbf{w}'\mathbf{F}'(t)\mathbf{x} + (1/2)\|\mathbf{H}(t)\mathbf{w}\|^2_{\mathbf{R}^{-1}(t)} \tag{26}$$

and consider the associated *canonical* differential equations

$$\left.\begin{aligned} d\mathbf{x}/dt &= \partial\mathcal{H}/\partial\mathbf{w}^5 = -\mathbf{F}'(t)\mathbf{x} + \mathbf{H}'(t)\mathbf{R}^{-1}(t)\mathbf{H}(t)\mathbf{w} \\ d\mathbf{w}/dt &= -\partial\mathcal{H}/\partial\mathbf{x} = \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}'(t)\mathbf{x} + \mathbf{F}(t)\mathbf{w} \end{aligned}\right\} \tag{27}$$

We denote the transition matrix of (27) by

$$\mathbf{\Theta}(t, t_0) = \begin{bmatrix} \mathbf{\Theta}_{11}(t, t_0) & \mathbf{\Theta}_{12}(t, t_0) \\ \mathbf{\Theta}_{21}(t, t_0) & \mathbf{\Theta}_{22}(t, t_0) \end{bmatrix} \tag{28}$$

[4] This is the number of distinct elements of the *symmetric* matrix $\mathbf{P}(t)$.

[5] The notation $\partial\mathcal{H}/\partial\mathbf{w}$ means the gradient of the scalar $\mathcal{H}$ with respect to the vector $\mathbf{w}$.

In Section 10 we shall prove

THEOREM 2. *The solution of* (IV) *for arbitrary nonnegative definite, symmetric* $\mathbf{P}_0$ *and all* $t \geqq t_0$ *can be represented by the formula*

$$\mathbf{\Pi}(t;\ \mathbf{P}_0, t_0) = [\mathbf{\Theta}_{21}(t, t_0) + \mathbf{\Theta}_{22}(t, t_0)\mathbf{P}_0]\cdot[\mathbf{\Theta}_{11}(t, t_0) + \mathbf{\Theta}_{12}(t, t_0)\mathbf{P}_0]^{-1} \tag{29}$$

Unless all matrices occurring in (27) are constant, this result simply replaces one difficult problem by another of similar difficulty, since only in the rarest cases can $\mathbf{\Theta}(t, t_0)$ be expressed in analytic form. Something has been accomplished, however, since we have shown that *the solution of nonconstant estimation problems involves precisely the same analytic difficulties as the solution of linear differential equations with variable coefficients.*

(8) *Existence of steady-state solution.* If the time-interval over which data are available is infinite, in other words, if $t_0 = -\infty$, Theorem 1 is not applicable without some further restriction.

For instance, if $\mathbf{H}(t) \equiv \mathbf{0}$, the variance of $\tilde{\mathbf{x}}$ is the same as the variance of $\mathbf{x}$; if the model (10–11) is unstable, then $\mathbf{x}(t)$ defined by (13) does not exist and the estimation problem is meaningless.

The following theorem, proved in Section 9, gives two sufficient conditions for the steady-state estimation problem to be meaningful. The first is the one assumed at the very beginning in the conventional Wiener theory. The second condition, which we introduce here for the first time, is much weaker and more "natural" than the first; moreover, it is almost a necessary condition as well.

THEOREM 3. *Denote the solutions of* (IV) *as in* (24). *Then the limit*

$$\lim_{t_0 \to -\infty} \mathbf{\Pi}(t;\ \mathbf{0}, t_0) = \bar{\mathbf{P}}(t) \tag{30}$$

exists for all t *and is a solution of* (IV) *if either*

(A_4) *the model* (10–11) *is uniformly asymptotically stable; or*

(A_4') *the model* (10–11) *is "completely observable"* [17], *that is, for all* t *there is some* $t_0(t) < t$ *such that the matrix*

$$\mathbf{M}(t_0, t) = \int_{t_0}^{t} \mathbf{\Phi}'(\tau, t)\mathbf{H}'(\tau)\mathbf{H}(\tau)\mathbf{\Phi}(\tau, t)d\tau \tag{31}$$

is positive definite. (*See* [21] *for the definition of uniform asymptotic stability.*)

Remarks. (*g*) $\bar{\mathbf{P}}(t)$ is the covariance matrix of the optimal error corresponding to the very special situation in which (i) an arbitrarily long record of past measurements is available, and (ii) the initial state $\mathbf{x}(t_0)$ was known exactly. When all matrices in (10–12) are constant, then so is also $\bar{\mathbf{P}}$—this is just the classical Wiener problem. In the constant case, $\bar{\mathbf{P}}$ is an equilibrium state of (IV) (i.e., for this choice of $\mathbf{P}$, the right-hand side of (IV) is zero). In general, $\bar{\mathbf{P}}(t)$ should be regarded as a moving equilibrium point of (IV), see Theorem 4 below.

(*h*) The matrix $\mathbf{M}(t_0, t)$ is well known in mathematical statistics. It is the *information matrix* in the sense of R. A. Fisher [20] corresponding to the special estimation problem when (i) $\mathbf{u}(t) \equiv \mathbf{0}$ and (ii) $\mathbf{v}(t)$ = gaussian with unit covariance matrix. In this case, the variance of any unbiased estimator $\mu(t)$ of $[\mathbf{x}^*, \mathbf{x}(t)]$ satisfies the well-known Cramér-Rao inequality [20]

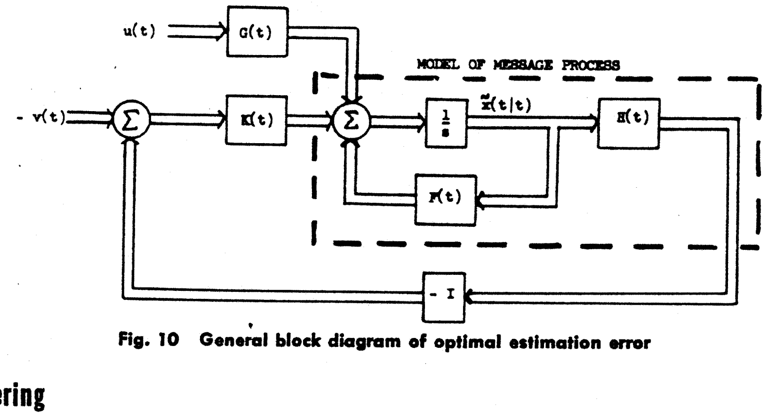

Fig. 10 General block diagram of optimal estimation error

$$\mathcal{E}[\mu(t) - \mathcal{E}\mu(t)]^2 \geqq \|\mathbf{x}^*\|^2_{\mathbf{M}^{-1}(t_0,\, t)} \tag{32}$$

Every costate $\mathbf{x}^*$ *has a minimum-variance unbiased estimator for which the equality sign holds in (32) if and only if* $\mathbf{M}$ *is positive definite.* This motivates the use of condition (A_4') in Theorem 3 and the term "completely observable."

(i) It can be shown [17] that in the constant case complete observability is equivalent to the easily verified condition:

$$\text{rank}[\mathbf{H}', \mathbf{F}'\mathbf{H}', \ldots, (\mathbf{F}')^{n-1}\mathbf{H}'] = n \tag{33}$$

where the square brackets denote a matrix with n rows and np columns.

(9) *Stability of the optimal filter.* It should be realized now that the *optimality* of the filter (I) does not at the same time guarantee its *stability.* The reader can easily check this by constructing an example (for instance, one in which (10–11) consists of two noninteracting systems). To establish weak sufficient conditions for stability entails some rather delicate mathematical technicalities which we shall bypass and state only the best final result currently available.

First, some additional definitions.

We say that the model (10–11) is *uniformly completely observable* if there exist fixed constants, α_1, α_2, and σ such that

$$\alpha_1\|\mathbf{x}^*\|^2 \leqq \|\mathbf{x}^*\|^2_{\mathbf{M}(t-\sigma,\, t)} \leqq \alpha_2\|\mathbf{x}^*\|^2 \quad \text{for all} \quad \mathbf{x}^* \text{ and } t.$$

Similarly, we say that a model is *completely controllable* [*uniformly completely controllable*] if the dual model is completely observable [uniformly completely observable]. For a discussion of these motions, the reader may refer to [17]. It should be noted that the property of "uniformity" is always true for constant systems.

We can now state the central theorem of the paper:

THEOREM 4. *Assume that the model of the message process is*

(A_4'') *uniformly completely observable;*
(A_5) *uniformly completely controllable;*
(A_6) $\alpha_3 \leqq \|\mathbf{Q}(t)\| \leqq \alpha_4, \quad \alpha_5 \leqq \|\mathbf{R}(t)\| \leqq \alpha_6$ *for all* t;
(A_7) $\|\mathbf{F}(t)\| \leqq \alpha_7$.

Then the following is true:

(i) *The optimal filter is uniformly asymptotically stable;*

(ii) *Every solution* $\mathbf{\Pi}(t;\ \mathbf{P}_0,\ t_0)$ *of the variance equation* (IV) *starting at a symmetric nonnegative matrix* $\mathbf{P}_0$ *converges to* $\bar{\mathbf{P}}(t)$ *(defined in Theorem 3) as* $t \to \infty$.

Remarks. (j) A filter which is not *uniformly* asymptotically stable may have an unbounded response to a bounded input [21]; the practical usefulness of such a filter is rather limited.

(k) Property (ii) in Theorem 4 is of central importance since it shows that the variance equation is a "stable" computational method that may be expected to be rather insensitive to roundoff errors.

(l) The speed of convergence of $\mathbf{P}_0(t)$ to $\bar{\mathbf{P}}(t)$ can be estimated quite effectively using the second method of Lyapunov; see [17].

(10) *Solution of the classical Wiener problem.* Theorems 3 and 4 have the following immediate corollary:

THEOREM 5. *Assume the hypotheses of Theorems 3 and 4 are satisfied and that* $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, $\mathbf{Q}$, $\mathbf{R}$, *are constants.*

Then, if $t_0 = -\infty$, *the solution of the estimation problem is obtained by setting the right-hand side of* (IV) *equal to zero and solving the resulting set of quadratic algebraic equations. That solution which is nonnegative definite is equal to* $\bar{\mathbf{P}}$.

To prove this, we observe that, by the assumption of constancy, $\bar{\mathbf{P}}(t)$ is a constant. By Theorem 4, all solutions of (IV) starting at nonnegative matrices converge to $\bar{\mathbf{P}}$. Hence, if a matrix $\mathbf{P}$ is found for which the right-hand side of (IV) vanishes and if this matrix is nonnegative definite, it must be identical with $\bar{\mathbf{P}}$. Note, however, that the procedure may fail if the conditions of Theorems 3 and 4 are not satisfied. See Example 4.

(11) *Solution of the Dual Problem.* For details, consult [17]. The only facts needed here are the following: The optimal control law is given by

$$\mathbf{u}^*(t^*) = -\mathbf{K}^*(t^*)\mathbf{x}(t^*) \tag{34}$$

where $\mathbf{K}^*(t^*)$ satisfies the duality relation

$$\mathbf{K}^*(t^*) = \mathbf{K}'(t) \tag{35}$$

and is to be determined by duality from formula (III). The value of the performance index (20) may be written in the form

$$\min_{\mathbf{u}^*} V(\mathbf{x}^*;\ t^*,\ t_0^*,\ \mathbf{u}^*) = \|\mathbf{x}^*\|^2_{\mathbf{\Pi}^*(t^*;\ \mathbf{x}^*,\, t^*_0)}$$

where $\mathbf{\Pi}^*(t^*;\ \mathbf{x}^*,\ t_0^*)$ is the solution of the dual of the variance equation (IV).

It should be carefully noted that the hypotheses of Theorem 4 are invariant under duality. Hence essentially the same theory covers both the estimation and the regular problem, as stated in Section 5.

The vector-matrix block diagram for the optimal regulator is shown in Fig. 11.

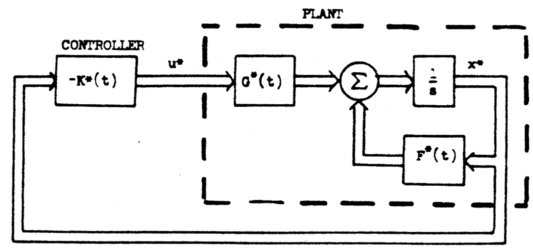

Fig. 11 General block diagram of optimal regulator

(12) *Computation of the covariance matrix for the message process.* To apply Theorem 1, it is necessary to determine cov $[\mathbf{x}(t_0), \mathbf{x}(t_0)]$. This may be specified as part of the problem statement as in Example 5. On the other hand, one might assume that the message model has reached steady state (see (A_3)), in which case from (13) and (12) we have that

$$\mathbf{S}(t) = \text{cov}\,[\mathbf{x}(t), \mathbf{x}(t)] = \int_{-\infty}^{t} \mathbf{\Phi}(t, \tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}'(\tau)\mathbf{\Phi}'(t, \tau)d\tau$$

provided the model (10) is asymptotically stable. Differentiating this expression with respect to t we obtain the following differential equation for $\mathbf{S}(t)$

$$d\mathbf{S}/dt = \mathbf{F}(t)\mathbf{S} + \mathbf{S}\mathbf{F}'(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}'(t) \tag{36}$$

This formula is analogous to the well-known lemma of Lyapunov [21] in evaluating the integrated square of a solution of a linear differential equation. In case of a constant system, (36) reduces to a system of linear algebraic equations.

8 Derivation of the Fundamental Equations

We first deduce the matrix form of the familiar Wiener-Hopf integral equation. Differentiating it with respect to time and then using (10–11), we obtain in a very simple way the fundamental equations of our theory.

Much cumbersome manipulation of integrals can be avoided by recognizing, as has been pointed out by Pugachev [27], that the Wiener-Hopf equation is a special case of a simple geometric principle: *orthogonal projection.*

Consider an abstract space $\mathfrak{X}$ such that an inner product (X, Y) is defined between any two elements X, Y of $\mathfrak{X}$. The norm is defined by $\|X\| = (X, X)^{1/2}$. Let $\mathfrak{U}$ be a subspace of $\mathfrak{X}$. We

$(10) \quad \frac{dx}{dt} = F(t)x + G(t)u$

Combining (10) and (I), we obtain the differential equation for the error of the optimal estimate: (10)-(I) ⇒

$$d\tilde{x}(t|t)/dt = [F(t) - K(t)H(t)]\tilde{x}(t|t) + G(t)u(t) - K(t)v(t) \quad \text{(II)}$$

To obtain an explicit expression for $K(t)$, we observe first that (39) implies that following identity in the interval $t_0 \leqq \sigma < t$:

$$\text{cov}\,[x(t), y(\sigma)] - \int_{t_0}^{t} A(t, \tau)\,\text{cov}\,[y(\tau), y(\sigma)]d\tau = A(t, \sigma)R(\sigma) \quad (39')$$

Since both sides of (39′) are continuous functions of σ, it is clear that equality holds also for $\sigma = t$. Therefore

$$K(t)R(t) = A(t, t)R(t) = \text{cov}[\tilde{x}(t|t), y(t)] = \text{cov}\,[\tilde{x}(t|t), x(t)]H'(t)$$

By (40), we have then

$$= \text{cov}\,[\tilde{x}(t|t), \tilde{x}(t|t)]H'(t) = P(t)H'(t)$$

Since $R(t)$ is assumed to be positive definite, it is invertible and therefore

$$K(t) = P(t)H'(t)R^{-1}(t) \quad \text{(III)}$$

We can now derive the variance equation. Let $\Psi(t, \tau)$ be the common transition matrix of (I) and (II). Then

$$P(t) - \Psi(t, t_0)P(t_0)\Psi'(t, t_0) = \mathcal{E}\int_{t_0}^{t} \Psi(t, \tau)[G(\tau)u(\tau) - K(\tau)v(\tau)]d\tau \times \int_{t_0}^{t} [u'(\sigma)G'(\sigma) - v'(\sigma)K'(\sigma)]\Psi'(t, \sigma)d\sigma$$

Using the fact that $u(t)$ and $v(t)$ are uncorrelated white noise, the integral simplifies to

$$= \int_{t_0}^{t} \Psi(t, \tau)[G(\tau)Q(\tau)G'(\tau) + K(\tau)R(\tau)K'(\tau)]\Psi'(t, \tau)d\tau$$

Differentiating with respect to t and using (III), we obtain after easy calculations the variance equation

$$dP/dt = F(t)P + PF'(t) - PH'(t)R^{-1}(t)H(t)P + G(t)Q(t)G'(t) \quad \text{(IV)}$$

Alternately, we could write

$$dP/dt = d\,\text{cov}\,[\tilde{x}, \tilde{x}]/dt = \text{cov}\,[d\tilde{x}/dt, \tilde{x}] + \text{cov}\,[\tilde{x}, d\tilde{x}/dt]$$

and evaluate the right-hand side by means of (II). A typical covariance matrix to be computed is

$$\text{cov}\,[\tilde{x}(t|t), u(t)] = \text{cov}\left[\int_{t_0}^{t} \Psi(t, \tau)[G(\tau)u(\tau) - K(\tau)v(\tau)]d\tau, u(t)\right] = (1/2)G(t)Q(t)$$

the factor $1/2$ following from properties of the δ-function.

To complete the derivations, we note that, if $t_1 > t$, then by (3)

$$x(t_1) = \Phi(t_1, t)x(t) + \int_{t}^{t_1} \Phi(t_1, \tau)u(\tau)d\tau$$

Since $u(\tau)$ for $t < \tau \leqq t_1$ is independent of $x(\tau)$ in the interval $t_0 \leqq \tau \leqq t$, it follows by (38) that the optimal estimator for the right-hand side above is 0. Hence

$$\hat{x}(t_1|t) = \Phi(t_1, t)\hat{x}(t|t) \quad (t_1 \geqq t) \quad \text{(V)}$$

The same conclusion does *not* follow if $t_1 < t$ because of lack of independence between $x(\tau)$ and $u(\tau)$.

The only point remaining in the proof of Theorem 1 is to determine the initial conditions for (IV). From (38) it is clear that

$$\hat{x}(t_0|t_0) = 0$$

Hence

$$P_0 = P(t_0) = \text{cov}[\tilde{x}(t_0|t_0), \tilde{x}(t_0|t_0)] = \text{cov}[x(t_0), x(t_0)]$$

In case of the conventional Wiener theory (see (A_3)), the last term is evaluated by means of (36).

This completes the proof of Theorem 1.

9 Outline of Proofs

Using the duality relations (16), all proofs can be reduced to those given for the regulator problem in [17].

(1) The fact that solutions of the variance equation exist for all $t \geqq t_0$ is proved in [17, Theorem (6.4)], using the fact that the variance of $x(t)$ must be finite in any finite interval $[t_0, t]$.

(2) Theorem 3 is proved by showing that there exists a particular estimate of finite but not necessarily minimum variance. Under (A_4'), this is proved in [17; Theorem (6.6)]. A trivial modification of this proof goes through also with assumption (A_4).

(3) Theorem 4 is proved in [17; Theorems (6.8), (6.10), (7.2)]. The stability of the optimal filter is proved by noting that the estimation error plays the role of a Lyapunov function. The stability of the variance equation is proved by exhibiting a Lyapunov function for P. This Lyapunov function in the simplest case is discussed briefly at the end of Example 1. While this theorem is true also in the nonconstant case, at present one must impose the somewhat restrictive conditions (A_4 – A_7).

10 Analytic Solution of the Variance Equation

Let $X(t)$, $W(t)$ be the (unique) matrix solution pair for (27) which satisfy the initial conditions

$$X(t_0) = I, \quad W(t_0) = P_0 \quad (47)$$

Then we have the following identity

$$W(t) = P(t)X(t), \quad t \geqq t_0 \quad (48)$$

which is easily verified by substituting (48) with (IV) into (27). On the other hand, in view of (47–48), we see immediately from the first set of equations (27) that $X(t)$ is the transition matrix of the differential equation

$$dx/dt = -F'(t)x + H'(t)R^{-1}(t)H(t)P(t)x$$

which is the adjoint of the differential equation (IV) of the optimal filter. Since the inverse of a transition matrix always exists, we can write

$$P(t) = W(t)X^{-1}(t), \quad t \geqq t_0 \quad (49)$$

This formula may not be valid for $t < t_0$, for then $P(t)$ may not exist!

Only trivial steps remain to complete the proof of Theorem 2.

11 Examples: Solution

Example 1. If $q_{11} > 0$ and $r_{11} > 0$, it is easily verified that the conditions of Theorems 3–4 are satisfied. After trivial substitutions in (III-IV) we obtain the expression for the optimal gain

$$k_{11}(t) = p_{11}(t)/r_{11} \quad (50)$$

and the variance equation

$$dp_{11}/dt = 2f_{11}p_{11} - p_{11}^2/r_{11} + q_{11} \quad (51)$$