Reprinted by permission from IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS
Vol. SMC-1, No. 2, April 1971
pp. 119-128

The Kalman–Bucy Filter as a True Time-Varying Wiener Filter

BRIAN D. O. ANDERSON, MEMBER, IEEE, AND JOHN B. MOORE, MEMBER, IEEE

***Abstract*—The notion is exploded that to build a Kalman–Bucy filter, one needs to know the whole structure of the signal generating process. It is shown that the filter is constructible knowing precisely those covariances required to construct a Wiener filter, and no more, and that the filter is independent of the particular models of the processes generating these covariances. Performance of the Kalman–Bucy filter does depend on the models, however. Results are also obtained for the smoothing problem.**

Manuscript received June 22, 1970; revised November 13, 1970. This work was supported by the Australian Research Grants Committee, the Australian–American Educational Foundation, and the Air Force Office of Scientific Research under Contract F44620-68-C-0023. Part of this work is based on research conducted by Dr. Anderson while at the Information and Control Sciences Center, Institute of Technology, Southern Methodist University, Dallas, Texas.
The authors are with the Department of Electrical Engineering, University of Newcastle, New South Wales 2308, Australia.

I. Introduction

TWO OF THE MOST important contributions in filtering theory have been the Wiener filter [1] and the Kalman–Bucy filter [2]. In Fig. 1 the usual situation where filtering may take place is illustrated. The various symbols have the following meanings: $y(\cdot)$ is a random process called the signal process; $n(\cdot)$ is a random process called the noise process; and $z(\cdot)$, given by $z(t) = y(t) + n(t)$, is a random process called the measurement process.

The Wiener or Kalman–Bucy filters are normally used when $y(\cdot)$, $n(\cdot)$, and $z(\cdot)$ are zero-mean Gaussian processes, $n(\cdot)$ is a white noise process, and $y(\cdot)$ is a process such that if a minimum variance estimate of $y(t)$ is constructible from $z(\tau)$, $\tau < t$, then the associated error variance is

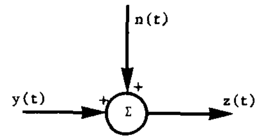

Fig. 1. Signal process $y(\cdot)$, noise process $n(\cdot)$, and measurement process $z(\cdot)$.

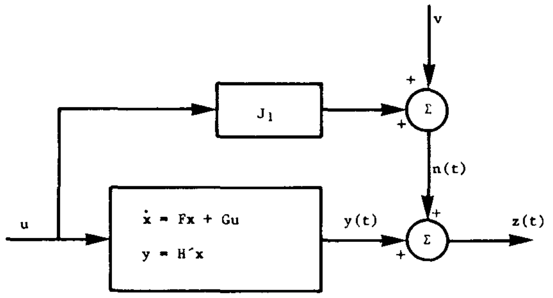

Fig. 2. Common filtering situation.

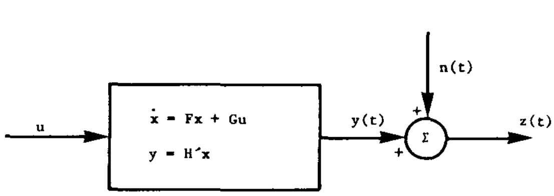

Fig. 3. Specialization of filtering situation of Fig. 2.

finite. This means, for example, that if $y(\cdot)$ is stationary, then the power spectral density of $y(\cdot)$ has to fall off at high frequencies at an appropriately fast rate; it also means that if $y(\cdot)$ is the output of a linear finite-dimensional system excited by white noise, then the system has no direct feedthrough term, so that each entry of y is a linear functional of the system state, rather than the system state and input.

Of course there are many possible variations of the situation just described which have been considered. However, this situation is probably the most common in which the Wiener or Kalman–Bucy filter will be applied. Moreover, it is the situation which most readily permits comparison of the differences between the two filters. Some of these major differences are as follows.

1) The Wiener filter may only be used when $y(\cdot)$ and $n(\cdot)$ are stationary[1] and when $z(\cdot)$ is available from time $-\infty$. In contrast the Kalman–Bucy filter can be used when $y(\cdot)$ and $n(\cdot)$ are nonstationary and when measurements $z(\cdot)$ are available commencing at some time t_0 with $t_0 > -\infty$ permitted.

2) The Kalman filter may only be used when the signal process $y(\cdot)$ is the output of a linear finite-dimensional system driven by white noise. In contrast the Wiener filter may be used irrespective of whether $y(\cdot)$ is the output of a specifically finite-dimensional linear system driven by white noise.

[1] More precisely, the solution of the resulting integral equation is greatly facilitated by assuming stationarity.

3) Apparently, the Kalman–Bucy filter can only be designed when knowledge is assumed of the finite-dimensional system generating the signal process $y(\cdot)$. In contrast all that is required to design a Wiener filter is knowledge of the quantities

$$E[y(t)y'(\tau)] + E[y(t)n'(\tau)], \qquad \text{for } t \geq \tau$$

$$E[z(t)z'(\tau)] = E[y(t)y'(\tau)] + E[y(t)n'(\tau)] + E[n(t)y'(\tau)] + E[n(t)n'(\tau)], \qquad \text{for } t \geq \tau.$$

There are two important special cases which we wish to note. The first arises when $y(t)$ is the output of a linear finite-dimensional system excited by white noise, and $n(t)$ is the sum of the two white-noise components, one derived from the system input $u(\cdot)$ and the other an independent process $v(\cdot)$ (see Fig. 2). In this case

$$E[n(t)y'(\tau)] = 0, \qquad \text{for } t > \tau \tag{1}$$

and so derivation of the Wiener filter requires knowledge of

$$E[y(t)y'(\tau)] + E[y(t)n'(\tau)], \qquad \text{for } t \geq \tau$$

$$E[n(t)n'(\tau)], \qquad \text{for } t \geq \tau.$$

Though knowledge of $E[y(t)y'(\tau)]$ and $E[y(t)n'(\tau)]$ separately is not required for filter design, it is required for the analysis of filter performance.

A second special case which arises is illustrated in Fig. 3. There the added white noise $n(\cdot)$ is independent of the signal process $y(\cdot)$, and thus

$$E[n(t)y'(\tau)] = 0, \qquad \text{for all } t,\tau. \tag{2}$$

Derivation of the Wiener filter then requires knowledge of

$$E[y(t)y'(\tau)], \qquad \text{for all } t \geq \tau$$

$$E[n(t)n'(\tau)], \qquad \text{for all } t \geq \tau.$$

For the remainder of this paper, we shall assume restriction (1) to be in force. At times we shall require the stronger restriction (2).

Point 3) would not constitute a real distinction between the two filters if one were able to pass from a prescribed $E[y(t)y'(\tau)] + E[y(t)n'(\tau)]$ to a finite-dimensional system with input white noise and output $y(\cdot)$ so that $y(\cdot)$ together with $n(\cdot)$ had the required statistical properties. Actually, this can be done under the constraint (2) for a scalar process $y(\cdot)$ (see [3], [4] for a procedure); unpublished work suggests that the scalar constraint can be removed. However, the procedure is computationally and conceptually difficult. Moreover, the linear finite-dimensional system deduced by the procedure is not uniquely specified. (The reader may recognize that the problem of devising a system which when driven by white noise has output covariance $E[y(t)y'(\tau)]$ is one of spectral factorization—known to be a difficult problem, and one with a nonunique solution.)

Examination of the Wiener filter derivation shows that at no stage is there a step corresponding to construction of a system with input white noise and output $y(t)$ satisfying the appropriate statistical constraints. Certainly, a spectral factorization of $E[z(t)z'(\tau)]$ is required, but such an operation is essentially carried out in the Kalman–Bucy filter derivation as well, but with knowledge of a system generating $E[y(t)y'(\tau)]$ assumed. (Note that because $z(t)$ contains white noise, the spectral factorization of $E[z(t)z'(\tau)]$ is more straightforward than the spectral factorization of $E[y(t)y'(\tau)]$, where $y(t)$ does not contain white noise [3], [4].)

The fact that a system generating $E[y(t)y'(\tau)]$ is not unique is perhaps disturbing. An initial study of the Kalman–Bucy filter might suggest that to each signal generating system, there would correspond a distinct filter. Yet, it is well known that the minimum-variance estimate of $y(t)$ produced by the Kalman–Bucy filter is unique and therefore presumably independent of the particular system generating the $y(\cdot)$ process. The rationale for asking the following questions should now be clear from the preceding remarks.

1) Can a Kalman–Bucy filter estimating $y(t)$, or even the signal process state, be derived assuming knowledge merely of

$$E[y(t)y'(\tau)] + E[y(t)n'(\tau)], \qquad \text{for } t \geq \tau$$
$$E[n(t)n'(\tau)], \qquad \text{for } t \geq \tau$$

rather than knowledge of a system generating a signal process $y(t)$ with the requisite statistical properties, and without construction of such a system by spectral factorization of $E[y(t)y'(\tau)]$?

2) Is the Kalman–Bucy filter independent of the particular system assumed to generate the signal process?

The answer to these questions is yes. This answer was in fact implicitly first given in [5] and in outline in [6], but these questions were not explicitly stated therein. Nor were the full ramifications of the answers investigated. Here we shall show how the Kalman–Bucy filter can be constructed by performing a spectral factorization of $E[z(t)z'(\tau)]$, this being in strict analogy with the Wiener theory. Knowledge of a system generating the $y(t)$ process is not assumed.

We also show that the filter, regarded as an estimator of $y(t)$, is independent of the particular system used to generate the signal process $y(t)$. (This is straightforwardly done from the Wiener–Hopf equation.) What is not quite so straightforwardly done is to show that the filter, when regarded as a filter estimating the state of the system generating the signal process $y(\cdot)$, is also independent of that system (provided we take some elementary precautions in setting up the coordinate basis of the state space to ensure uniqueness of the basis). The error variance in the state estimate depends, however, on the particular signal process model.

Section II contains definitions and preliminaries. Section III establishes that the Kalman–Bucy filter, regarded not just as a device for estimating $y(t)$, but also as a device for estimating the state of a system generating $y(t)$, is independent of the particular generating system. Section IV derives the Kalman–Bucy filter, assuming no knowledge of a system generating the signal process, while Section V verifies by direct calculation that the filter as obtained in Section IV is the same as that obtained by the calculation of [2], assuming availability of the signal process model. Section VI notes the error variance property mentioned briefly in the preceding, and Section VII considers the smoothing problem. The optimal smoother for the signal process $y(t)$ turns out to be independent of the system generating $y(t)$ for a subclass of such systems, but the optimal smoother for the state of the signal model process depends on the particular signal model process.

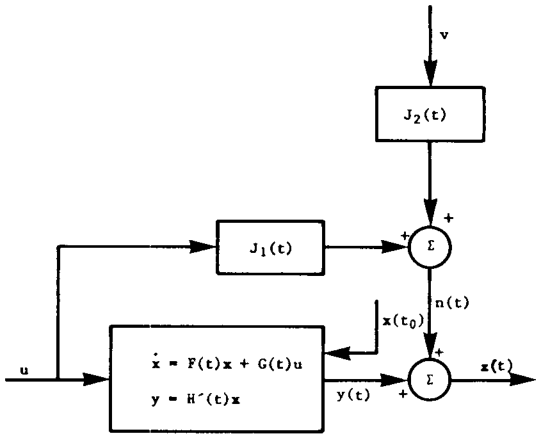

Fig. 4. General signal and noise model.

II. Definitions and Preliminary Analysis

Consider the arrangement of Fig. 4. Both $u(\cdot)$ and $v(\cdot)$ are independent vector random processes which are Gaussian, zero mean, and white with covariance

$$E[u(t)u'(\tau)] = E[v(t)v'(\tau)] = I\,\delta(t - \tau). \tag{3}$$

The initial state vector of the system $x(t_0)$ is a Gaussian random variable, assumed zero mean for convenience though this assumption is unnecessary. Also, we require $x(t_0)$ to be independent of $u(\cdot)$ and $v(\cdot)$ and such that

$$E[x(t_0)x'(t_0)] = P_0. \tag{4}$$

Let $P(t)$ be the solution of

$$\dot{P} = PF' + FP + GG', \qquad P(t_0) = P_0. \tag{5}$$

It is not hard to see that

$$E[x(t)x'(t)] = P(t) \tag{6}$$

and, in fact, one may readily verify the following equalities:

$$E[y(t)y'(\tau)] = H'(t)\Phi(t,\tau)P(\tau)H(\tau)1(t - \tau) + H'(t)P(t)\Phi'(\tau,t)H(\tau)1(\tau - t) \tag{7}$$

$$E[y(t)n'(\tau)] = H'(t)\Phi(t,\tau)G(\tau)J_1'(\tau)1(t - \tau) \tag{8}$$

$$E[n(t)n'(\tau)] = [J_1(t)J_1'(t) + J_2'(t)J_2(t)]\delta(t - \tau) \tag{9}$$

$$E[z(t)z'(\tau)] = R(t)\delta(t - \tau) + H'(t)\Phi(t,\tau)L(\tau)1(t - \tau) + L'(t)\Phi'(\tau,t)H(\tau)1(\tau - t) \tag{10}$$

where $\Phi(\cdot,\cdot)$ is the transition matrix associated with $\dot{x} = F(t)x$, the symbol $1(t - \tau)$ denotes the unit step function, zero for $t < \tau$, unity for $t > \tau$, and the matrices

$R(t)$ and $L(t)$ are defined by

$$R(t) = J_1(t)J_1'(t) + J_2(t)J_2'(t) \tag{11}$$

$$L(t) = P(t)H(t) + G(t)J_1'(t). \tag{12}$$

The scheme of Fig. 4 is typical of the process and noise models for which Kalman filters are constructed. In [7] a minor extension of the models of [2] was made to consider process and noise models described by

$$\begin{aligned} \dot{x} &= F(t)x(t) + G(t)u(t) \\ y(t) &= H'(t)x(t) \\ z(t) &= H'(t)x(t) + n(t). \end{aligned} \tag{13}$$

As in the preceding, $x(t_0)$, $u(\cdot)$, and $n(\cdot)$ are Gaussian, with zero mean, $x(t_0)$ is independent of $u(\cdot)$ and $n(\cdot)$, and

$$\begin{aligned} E[x(t_0)x(t_0)] &= P_0 \\ E[u(t)u'(\tau)] &= I\delta(t - \tau) \\ E[n(t)n'(\tau)] &= R(t)\delta(t - \tau). \end{aligned} \tag{14}$$

Actually, $E[u(t)u'(\tau)] = Q(t)\delta(t - \tau)$ is permitted, but $Q(\cdot)$ can be absorbed in $G(\cdot)$. Instead of $u(\cdot)$ and $n(\cdot)$ being independent, we have

$$E[u(t)n'(\tau)] = J_3(t)\delta(t - \tau). \tag{15}$$

The point to note is that there is now no direct feedthrough from u to z, but the additive noise at the output is correlated with the input noise. This arrangement is equivalent to one of the sort described by Fig. 4, where $u(\cdot)$ and $v(\cdot)$ are independent, and the equivalence is defined (as direct calculation shows) by the equations

$$n(t) = J_1u(t) + J_2v(t) \tag{16}$$

with

$$J_1(t) = J_3'(t) \tag{17}$$

and J_2 any matrix such that

$$J_2(t)J_2'(t) = R(t) - J_1(t)J_1'(t). \tag{18}$$

(Such a matrix is guaranteed to exist.)

In accordance with the remarks made in the previous section, it follows that the problem of estimating $y(t)$, given $z(\tau)$, $t_0 \leq \tau < t$, for any system of the form of Fig. 4, has a parallel with the Wiener filter class of problems if we assume knowledge of

$$E[y(t)y'(\tau)] + E[y(t)n'(\tau)] = H'(t)\Phi(t,\tau)L(\tau), \qquad \text{for } t \geq \tau \tag{19}$$

and if

$$E[n(t)n'(\tau)] = R(t)\delta(t - \tau), \qquad \text{for } t \geq \tau. \tag{20}$$

The matrices $H(\cdot)$, $\Phi(\cdot,\cdot)$, $L(\cdot)$, and $R(\cdot)$ do not contain sufficient information to identify G, J_1, J_2, and P_0, quantities which would all be required were a Kalman–Bucy filter to be designed by the usual procedure. In fact there are an infinity of G, J_1, J_2, and P_0 which will lead to the same L and R as in the preceding.

III. Invariance of Kalman–Bucy Filter with Respect to Signal Model

We suppose in this section that we have noisy measurements

$$z(t) = y(t) + n(t) \tag{21}$$

of a signal process $y(t)$. We suppose moreover that

$$E[y(t)n'(\tau)] = 0, \qquad t < \tau \tag{22}$$

and that for $t \geq \tau$, we have the following statistical information:

$$\begin{aligned} E[y(t)z'(\tau)] &= E[y(t)y'(\tau)] + E[y(t)n'(\tau)] \\ &= H'(t)\Phi(t,\tau)L(\tau) \end{aligned} \tag{23}$$

$$E[n(t)n'(\tau)] = R(t)\delta(t - \tau). \tag{24}$$

We shall demand the following further condition.

Assumption 1: $E[z(t)z'(\tau)]$ is positive definite for all $t,\tau \geq t_0$, and $R(t)$ is positive definite for all t.

Note that a sufficient condition for this assumption to hold is that $z(t)$ result from an arrangement as shown in Fig. 4 with $J_1(t)J_1'(t) + J_2(t)J_2'(t)$ nonsingular for all t. Using the Wiener–Hopf equation, we shall establish a result, almost intuitively obvious, concerning the estimation of $y(t)$. (A minor variant of this result appears in [10].)

Theorem 1

Consider the situation described by (21)–(24) and the associated remarks and suppose that Assumption 1 is in force. Then the linear system generating a minimum variance unbiased estimate of $y(t)$ from $z(\tau)$, $t_0 \leq \tau < t$, is independent of the particular model generating the signal process $y(\cdot)$.

Proof: According to the Wiener–Hopf equation, the impulse response $A_y(t,\tau)$ of the optimum or minimum-variance unbiased filter is given by

$$E[y(t)z'(\tau)] = \int_{t_0}^{t} A_y(t,\sigma)E[z(\sigma)z'(\tau)]\,d\sigma, \qquad t > \tau. \tag{25}$$

The quantities $E[y(t)z'(\tau)]$ and $E[z(\sigma)z'(\tau)]$ are known and are independent of the particular generating model, while Assumption 1 guarantees by a basic result of integral equation theory [8] that $A_y(\cdot,\cdot)$ is unique.

Far more interesting is the next result, concerning state estimation. This result states that no matter what system of the form of Fig. 4 generates the quantities (23) and (24), the Kalman–Bucy filter is the same. Before stating the result, however, it is necessary to pin down precisely the coordinate basis for the state space of the generating systems. Given the expression on the left of (23), the quantities $H(\cdot)$, $\Phi(\cdot,\cdot)$, or $F(\cdot)$ and $L(\cdot)$ are not uniquely determined. However, procedures of linear system theory are available which will generate a set of such quantities. The choice of any one set of $H(\cdot)$, $\Phi(\cdot,\cdot)$, and $L(\cdot)$ is tantamount to fixing a state-space coordinate basis, in this case the coordinate basis of any system generating the signal process

$y(\cdot)$. Since it would be futile to conceive of state estimation without a definite coordinate basis in mind, we make the following assumption.

Assumption 2: Quantities $H(\cdot)$, $\Phi(\cdot,\cdot)$, and $L(\cdot)$, with $\Phi(\cdot,\cdot)$ a transition matrix, are known which satisfy (23). Moreover, if $L_1(\cdot)$ and $L_2(\cdot)$ are such that

$$H'(t)\Phi(t,\tau)L_1(\tau) = H'(t)\Phi(t,\tau)L_2(\tau)$$

for all t and τ, then $L_1(\tau) = L_2(\tau)$, for all τ.

The last part of this assumption says that once $\Phi(\cdot,\cdot)$ and $H(\cdot)$ or $F(\cdot)$ and $H(\cdot)$ are known, $L(\cdot)$ is uniquely determined. It also amounts to demanding complete observability of any model of the signal process with the specified $F(\cdot)$ and $H(\cdot)$.

Theorem 2

Consider the situation described by (21)–(24) and the associated remarks, and suppose that Assumptions 1 and 2 are in force. Let $x(t)$ be the state at time t of a linear system generating the signal process $y(\cdot)$, with equations of the form

$$\dot{x} = F(t)x + G(t)u, \qquad E[x(t_0)x'(t_0)] = P_0$$
$$y = H'(t)x. \tag{26}$$

Then the linear system generating a minimum variance unbiased estimate $x_e(t)$ of $x(t)$, given the measurements $z(\tau)$, $t_0 \leq \tau < t$, is independent of the particular system (among those with the same $F(\cdot)$ and $H(\cdot)$) which is assumed to generate $y(t)$ and depends purely on $H(\cdot),\Phi(\cdot,\cdot)$ (or equivalently, $F(\cdot)$), $L(\cdot)$ and $R(\cdot)$.

Proof: The impulse response of the minimum-variance unbiased filter $A_x(t,\tau)$ is given by the Wiener–Hopf equation

$$E[x(t)z'(\tau)] = \int_{t_0}^{t} A_x(t,\sigma)E[z(\sigma)z'(\tau)]\,d\sigma, \qquad t > \tau. \tag{27}$$

Assumption 1 again guarantees existence and uniqueness of a solution to this integral equation, at least if we have $E[x(t)z'(\tau)]$ available, and this quantity has no delta function terms.

Let us now compute $E[x(t)z'(\tau)]$, for $t > \tau$, and thereby show that it is independent of the particular signal model and is determinable from $H(\cdot)$, $F(\cdot)$, $L(\cdot)$, and $R(\cdot)$. This will establish that the same filter is used for all signal process models of the allowed type.

For the system depicted in Fig. 4, it is not hard to calculate that

$$E[x(t)u'(\tau)] = \Phi(t,\tau)G(\tau), \qquad \text{for } t > \tau. \tag{28}$$

In Section II, we also computed $E[x(t)x'(t)]$ as follows:

$$E[x(t)x'(t)] = P(t) \tag{29}$$

where

$$\dot{P} = PF' + FP + GG', \qquad P(t_0) = P_0. \tag{30}$$

It is not hard to verify then that

$$E[x(t)x'(\tau)] = \Phi(t,\tau)P(\tau), \qquad \text{for } t > \tau. \tag{31}$$

Now we can use (28) and (31):

$$\begin{aligned} E[x(t)z'(\tau)] &= E[x(t)y'(\tau)] + E[x(t)n'(\tau)] \\ &= E[x(t)x'(\tau)]H(\tau) + E[x(t)u'(\tau)]J_1'(\tau) \\ &= \Phi(t,\tau)[P(\tau)H(\tau) + G(\tau)J_1'(\tau)]. \end{aligned}$$

We recall now from Section II that the quantity $L(\cdot)$ is defined by

$$L(\tau) = P(\tau)H(\tau) + G(\tau)J_1'(\tau) \tag{32}$$

so that

$$E[x(t)z'(\tau)] = \Phi(t,\tau)L(\tau), \qquad \text{for } t > \tau. \tag{33}$$

The desired independence is immediate. Notice that (33) implies by left multiplication by $H'(t)$ that

$$E[y(t)z'(\tau)] = H'(t)\Phi(t,\tau)L(\tau)$$

which is the same as (23). Development of (33) from (23), however, seems no easier than development along the lines we have just followed.

It is well known that a minimum variance unbiased estimate $y_e(t)$ of $y(t)$ is provided by $H'(t)x_e(t)$, where $x_e(t)$ is the minimum variance unbiased estimate of $x(t)$. Therefore, given a state estimator, a signal estimator is immediately constructible. One important point should be noted, however; this is that the estimate $y_e(t)$ can be obtained from $x_e(t)$ no matter what the coordinate basis for the state space is. In other words if the two triples $F_1(\cdot),H_1(\cdot),L_1(\cdot)$ and $F_2(\cdot),H_2(\cdot),L_2(\cdot)$ are such that

$$H_1'(t)\Phi_1(t,\tau)L_1(\tau) = H_2'(t)\Phi_2(t,\tau)L_2(\tau)$$

for all t and τ, with $\Phi_i(\cdot,\cdot)$ the transition matrix of $F_i(\cdot)$, $i = 1,2$, and if an appropriately modified version of Assumption 2 holds, y_e can be derived as $H_1'x_{1e}$, or equally well as $H_2'x_{2e}$. Here, for $i = 1,2$, x_{ie} is the state estimate with F_i and H_i defining the coordinate basis of the state space of the class of signal models.

IV. Filter Computation Without Signal Process Model

It is clear that with the aid of Theorem 2, we could conceive of a filter calculation as consisting of two parts; first, the calculation of any model whatsoever for the signal process, and second, the computation of a filter by the procedures of [2]. As we have noted earlier, the calculation of a model for the signal process, at least from $E[y(t)y'(\tau)]$, is a very difficult task. Moreover, the design of a Wiener filter requires spectral factorization merely of $E[z(t)z'(\tau)]$ and not of $E[y(t)y'(\tau)]$. The Wiener filter situation gives a clue as to what we should do. We shall proceed as follows.

1) We shall perform a spectral factorization of $E[z(t)z'(\tau)]$, i.e., we shall obtain a linear system such that with input white noise, the output covariance is $E[z(t)z'(\tau)]$. This system will have direct feedthrough between input and output, but, and this is important, it will be a member of the general class depicted in Fig. 4.

2) We shall observe that the problem of determining the state of the system mentioned in 1) can, in this special case, be solved immediately, the state of the system being recoverable with no error by feeding $z(\cdot)$ into a linear finite-dimensional system whose output is $x(t)$. Since the optimal filter is known to be unique, this second system must be, according to Theorem 2, the optimal filter for any system generating the signal process and possessing the required $F(\cdot)$ and $H(\cdot)$.

Theorem 3 (Spectral Factorization of $E[z(t)z'(\tau)]$)

Given a positive definite covariance $E[z(t)z'(\tau)]$ of the form [3]

$$E[z(t)z'(\tau)] = R(t)\delta(t-\tau) + H'(t)\Phi(t,\tau)L(\tau)1(t-\tau) + L'(t)\Phi'(\tau,t)H(\tau)1(\tau-t) \quad (34)$$

with $R(t)$ positive definite and $t,\tau \geq t_0$, there exists a nonnegative definite symmetric solution to the following differential equation, for all $t \geq t_0$:

$$\dot{P}_g(t) = P_g(t)F'(t) + F(t)P_g(t) + [P_g(t)H(t) - L(t)]R^{-1}(t)[P_g(t)H(t) - L(t)]'$$
$$P_g(t_0) = 0. \quad (35)$$

Moreover, if $E[u(t)u'(\tau)] = I\delta(t-\tau)$, a process $z(t)$ of covariance $E[z(t)z'(\tau)]$ may be obtained as the output of the following system:

$$\dot{x}_g = F(t)x_g(t) + [L(t) - P_g(t)H(t)]R^{-1/2}(t)u(t)$$
$$x(t_0) = 0$$
$$z(t) = H'(t)x(t) + R^{1/2}(t)u(t). \quad (36)$$

Actually, all claims save for the existence of $P_g(t)$ may be established using the results of Section II. Notice that the system of (36), shown in Fig. 5, is a special case of the class of systems depicted by Fig. 4. In particular the initial state covariance $E[x(t_0)x'(t_0)]$, earlier termed $P(t_0)$, is now zero, as is $J_2(t)$, for all t.

Estimation of $x_g(t)$

The fact that there is no white noise in $z(t)$ which is independent of $u(t)$ makes the problem of estimating $x_g(t)$ a simple problem of finding the inverse of a system. We claim that the scheme of Fig. 6, drawn in several different formats, will generate a zero error estimate $x_e(t)$ of $x_g(t)$. To see this, observe from (36) and

$$\dot{x}_e(t) = [F - (L - P_gH)R^{-1}H']x_e + (L - P_gH)R^{-1}z$$
$$x_e(t_0) = 0 \quad (37)$$

that

$$\dot{x}_e(t) - \dot{x}_g(t) = [F - (L - P_gH)R^{-1}H'][x_e(t) - x_g(t)].$$

Since also $x_e(t_0) = x_g(t_0) = 0$, it follows that $x_e(t) = x_g(t)$, for all t.

An estimate of $y_g(t) = H'x_g(t)$ is of course obtained, with zero error, as $H'x_e(t)$. Combining the preceding result with Theorem 2, we obtain the following theorem.

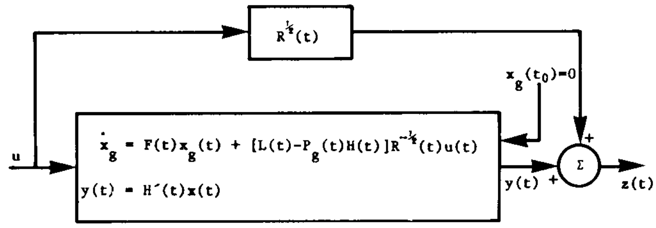

Fig. 5. System generating process $z(\cdot)$.

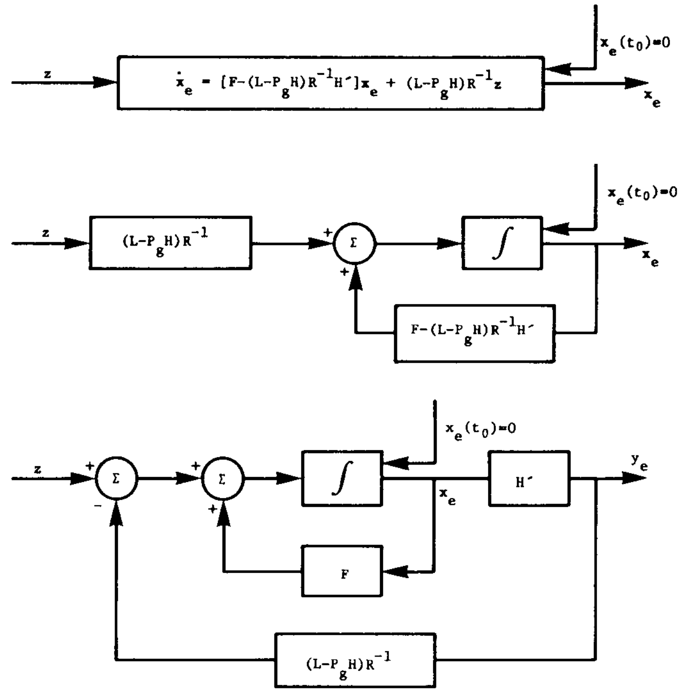

Fig. 6. Optimal filter for system of Fig. 5 drawn in different formats; also Kalman–Bucy filter for class of signal process models.

Theorem 4

Consider the situation described by (21)–(24) and the associated remarks, and suppose that Assumptions 1 and 2 are in force. Let $x(t)$ be the state at time t of any linear system generating the signal process $y(\cdot)$ with equations of the form

$$\dot{x} = F(t)x + G(t)u, \qquad E[x(t_0)x'(t_0)] = P_0$$
$$y = H'(t)x. \quad (38)$$

Then a minimum-variance unbiased estimate $x_e(t)$ of $x(t)$ is obtained via the scheme shown in Fig. 6 and described by (37), where the matrix $P_g(t)$ is obtained from (35). An unbiased minimum-variance estimate of $y(t)$ is provided by $H'(t)x_e(t)$.

We reiterate that only F, H, L, and R are required for the filter computation.

Notice that there are really two important properties required of the spectral factor of $E[z(t)z'(\tau)]$ in order that it be invertible. First, it must have a direct feedthrough component, so that construction of an inverse is possible. Second, the initial state covariance must be zero. The reason is as follows. If the system of Fig. 6 is to act as an exact state estimator, it is essential that the initial states $x_g(t_0)$ and $x_e(t_0)$ be matched at t_0. If $x_g(t_0)$ has a known deterministic value, this is straightforward. But in our case, $x_g(t_0)$ is a Gaussian random variable. The only way we can ensure that $x_e(t_0) = x_g(t_0)$ is that $x_g(t_0)$ actually be zero, i.e., a Gaussian random variable of mean and covariance zero. If a spectral factor of $E[z(t)z'(\tau)]$ is derived with $E[x_g(t_0)x_g'(t_0)]$ nonzero, there is no way of setting $E[x_e(t_0)x_e'(t_0)]$ correctly. This point is also discussed in [3] in connection with whitening filter design.

V. Rapprochement with Standard Filter Derivation: Example

In this section we first wish to check by direct calculation that the filter described in Theorem 4 is the same as that computed by standard procedures. We shall do this for the case when a system model is given as

$$\dot{x} = F(t)x(t) + G(t)u(t)$$
$$z = H'(t)x(t) + n(t) \qquad (39)$$

with $x(t_0)$, $u(\cdot)$, $n(\cdot)$ Gaussian, zero mean, and mutually independent, and such that

$$E[u(t)u'(\tau)] = I\delta(t - \tau)$$
$$E[n(t)n'(\tau)] = R(t)\delta(t - \tau)$$
$$E[x(t_0)x'(t_0)] = P_0. \qquad (40)$$

The Kalman–Bucy filter is defined (as we know from [2]) using the solution $P_k(\cdot)$ of

$$P_k = P_kF' + FP_k - P_kHR^{-1}H'P_k + GG'$$
$$P_k(t_0) = P(t_0). \qquad (41)$$

It is

$$\dot{x}_e(t) = (F - P_kHR^{-1}H')x_e(t) + P_kHR^{-1}z(t)$$
$$x_e(t_0) = 0. \qquad (42)$$

Comparison with (37), claimed in Theorem 4 to be the filter equation, shows that (42) and (37) will be the same if and only if

$$(L - P_gH)R^{-1} = P_kHR^{-1}. \qquad (43)$$

We shall now aim to establish this. First, we must establish what $L(\cdot)$ is, using the model (34). We recall that $L(\cdot)$ is given by

$$L = PH \qquad (44)$$

where

$$\dot{P} = PF' + FP + GG', \qquad P(t_0) = P_0. \qquad (45)$$

(Here, $P(t) = E[x(t)x'(t)]$.)

Equation (43) will be established if it is true that

$$P - P_g = P_k. \qquad (46)$$

But from the defining equation (35) for P_g and (44), we have

$$\dot{P}_g = P_gF' + FP_g + (P_g - P)HR^{-1}H'(P_g - P)$$

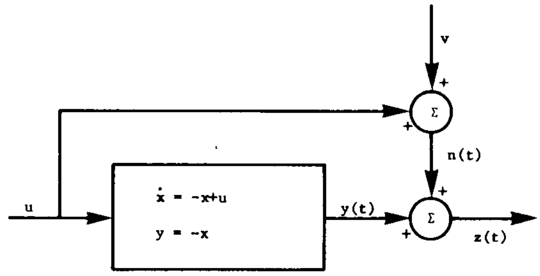

Fig. 7. Example of Section V.

which, when combined with (45), yields

$$\frac{d}{dt}(P - P_g) = (P - P_g)F' + F(P - P_g)$$
$$- (P - P_g)HR^{-1}H'(P - P_g) + GG' \qquad (47)$$

and the initial conditions on P and P_g imply $P(t_0) - P_g(t_0) = P_0$. Comparison of (47) with (41), including the initial conditions, establishes (46).

Let us now consider an example where we are required to filter the signal component of a covariance of the form

$$E[z(t)z(\tau)] = 2\delta(t - \tau) - \tfrac{1}{2}\exp[-(t - \tau)]1(t - \tau)$$
$$- \tfrac{1}{2}\exp[-(\tau - t)]1(\tau - t).$$

We are given that

$$z(t) = y(t) + n(t)$$

with $n(t)$ consisting of white noise, $y(t)$ containing no white noise, and

$$E[y(t)n'(\tau)] = 0, \qquad \text{for } \tau > t.$$

This example has been chosen for the reason that there is no signal and noise model of the form of Fig. 3 which will generate the process $z(t)$. In other words, one cannot even force this problem into one where standard Kalman–Bucy filtering could be applied. Certainly, though the covariance is generatable by a scheme of the form of Fig. 4, in fact, Fig. 7 shows one such scheme. Here $t_0 = -\infty$, and u and v are of course independent zero mean Gaussian white noise processes, with covariance $\delta(t - \tau)$. It can be verified by direct calculation that the process of $z(t)$ of Fig. 7 has the requisite covariance.

Perhaps one could now conceive of estimating $y(t)$ by forming a Kalman filter for the process of Fig. 7, although one must still have had the insight to obtain Fig. 7 in the first place. But it is easier to work directly from the covariance of $z(\cdot)$. In terms of the notation used hitherto, we have

$$R(t) = 2, \qquad F(t) = -1, \qquad H(t) = -1, \qquad L(t) = \tfrac{1}{2}$$

and thus

$$\dot{P}_g = -2P_g + [P_g + \tfrac{1}{2}]\tfrac{1}{2}[P_g + \tfrac{1}{2}]$$

with initial condition $\lim_{t_0 \to -\infty} P_g(t_0) = 0$. The solution is

$$\dot{P}_g(t) = 3 - \frac{\sqrt{35}}{2}$$

so that the optimal filter is

$$\dot{x}_e = \left\{\frac{3}{4} - \frac{\sqrt{35}}{4}\right\} x_e + \left\{\frac{7}{4} - \frac{\sqrt{35}}{3}\right\} 2, \qquad \lim_{t_0 \to -\infty} x_e(t_0) = 0.$$

Kalman and Bucy [2] do not actually explain how a Kalman–Bucy filter may be derived for the scheme of Fig. 7. By redrawing the scheme to be in the form of Fig. 3, but with $u(t)$ and $n(t)$ correlated, material in [7] can be applied to calculate the filter. The calculations are a good deal longer.

VI. Error Variances

In this section we wish to note quantitatively how the error variance in estimating the state and output of a linear system depends both on the model and on the statistics required for filter derivation. We shall see that even though the same filter may be used for an infinity of different models, the performance of the filter will generally differ from model to model. We discuss first the error in estimating the output $y(t)$ of a system as shown in Fig. 4.

Theorem 5

Consider the situation described by (21)–(24) with Assumptions 1 and 2 holding. Let $y_e(t)$ be the optimal estimate of $y(t)$ obtained as described in Theorem 4. Then $E[y(t) - y_e(t)]^2$ depends only on $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$.

Proof: First note that the optimal filter for $y(\cdot)$ depends on $E[n(t)n'(\tau)]$ and on $E[y(t)y'(\tau)] + E[y(t)n'(\tau)]$, for $t > \tau$, but not on the summands separately. Note also that there will be an infinity of systems which yield the same $E[y(t)n'(\tau)]$, though two systems yielding the same $E[y(t)y'(\tau)] + E[y(t)n'(\tau)]$ will not necessarily yield the same values of each summand.

The easiest way to prove the theorem is to appeal to the result that [2]

$$E\{[x(t) - x_e(t)][x(t) - x_e(t)]\} = P_k(t) \tag{48}$$

where $P_k(t)$ is defined by (41) or by

$$P_k(t) = P(t) - P_g(t) \tag{49}$$

which is (46) repeated, and $P(t)$ and $P_g(t)$ are given by (5) and (35), respectively. From (48), it follows that

$$\begin{aligned} E\{[y(t) - y_e(t)][y(t) - y_e(t)]'\} &= H'(t)P_k(t)H(t) \\ &= H'(t)P(t)H(t) \\ &\quad - H(t)P_g(t)H(t). \end{aligned} \tag{50}$$

From (6), repeated as

$$E[x(t)x'(t)] = P(t)$$

it follows that

$$H'(t)P(t)H(t) = E[y(t)y'(t)].$$

The matrices $P_g(t)$ and $H(t)$ are independent of the particular signal model process, i.e., they depend purely on $E[y(t)y'(\tau)] + E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$. Therefore, both terms on the right side of (50) can be considered as depending on $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$. This proves the theorem.

When we turn to the error variance associated with the state estimate, the conclusion we reach is that there can be no connection between the error variance associated with different signal model processes. Construction of a few examples will verify the intuitively obvious fact that the covariance of the state of the signal process model $P(t)$ will vary from model to model. Since $P_g(t)$ is independent of the model, it follows that the error variance $P_k(t) = P(t) - P_g(t)$ again depends on the particular signal process model.

The special case arising when $E[y(t)n'(\tau)] = 0$, for all t and τ, is worthy of consideration. For this constraint, it follows that $E\{[y(t) - y_e(t)][y(t) - y_e(t)]'\}$ depends purely on $E[y(t)y'(\tau)]$ and $E[n(t)n'(\tau)]$. However, $E\{[x(t) - x_e(t)][x(t) - x_e(t)]'\}$ still depends on the actual signal model; in other words, $H'PH$ and $H'P_kH$ are independent of the model, but P and P_k are not. With $E[y(t)n'(\tau)] \equiv 0$, it is possible to show [9] that there is one particular generating system—actually causally invertible—such that the associated $P_k(t)$ is minimal for all t, i.e., if $\bar{P}_k(t)$ is associated with any other signal process model such that $E[y(t)n'(\tau)] \equiv 0$, then $\bar{P}_k(t) - P_k(t)$ is nonnegative definite for all t. Of course, among the set of signal models for which $E[y(t)n'(\tau)]$ is not restricted to being zero, there is one for which the associated $P_k(t)$ is identically zero. This system was defined in our derivation of the filter equation based on spectral factorization of $E[z(t)z'(\tau)]$.

VII. Smoothing Problem

In this section we wish to note precisely what data is required to obtain a minimum-variance unbiased estimate of $y(t)$ given measurements $z(\tau)$, for $t_0 \leq \tau \leq t + \Delta$, where $\Delta > 0$. We shall note that slightly more information is required than in the case of the filtering problem. We shall also study the problem of finding an estimate of the state of a signal process model $x(t)$. We shall find that the filter producing an estimate of $x(t)$ depends on the particular process model.

Suppose that the smoothed estimate of $y(t)$ becomes available at time $t + \Delta$ by inserting the $z(\cdot)$ process into a linear system of impulse response $A_y(t + \Delta,\tau)$. Then

$$E[y(t)z'(\tau)] = \int_{t_0}^{t+\Delta} A_y(t + \Delta,\sigma)E[z(\sigma)z'(\tau)]\, d\sigma, \qquad t_0 \leq \tau < t + \Delta. \tag{51}$$

This equation has a unique solution under Assumption 1, assuming availability of $E[y(t)z'(\tau)]$ for the appropriate values of t and τ. For $\tau \leq t$,

$$E[y(t)z'(\tau)] = E[y(t)y'(\tau)] + E[y(t)n'(\tau)] \tag{52}$$

while for $\tau > t$, assuming the fundamental constraint of (1), viz., $E[y(t)n'(\tau)] = 0$, for $t > \tau$, we have the following equation:

$$E[y(t)z'(\tau)] = E[y(t)y'(\tau)]. \quad (53)$$

In view of the symmetry of $E[y(t)y'(\tau)]$, knowledge of this quantity for $t < \tau < t + \Delta$ is equivalent to knowledge of this quantity for $\tau < t < \tau + \Delta$, or simply $\tau < t$ if Δ is arbitrary. Therefore, knowledge of the quantities on the left side of (52) and (53)—essential for solvability of (51)—is equivalent to knowledge, as separate quantities, of $E[y(t)y'(\tau)]$ and $E[y(t)n'(\tau)]$.

If two signal process models have the same value of $E[y(t)y'(\tau)] + E[y(t)n'(\tau)]$ but different values of, say, $E[y(t)y'(\tau)]$, though the optimal filters will be the same, the optimal smoothers will not. Summing up, we state the following theorem.

Theorem 6

Consider the situation described by (21) and (22) and suppose that Assumption 1 is in force. Then the linear system generating a minimum variance unbiased estimate of $y(t)$ from $z(\tau)$, $t_0 \leq \tau < t + \Delta$, depends on $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$, but is otherwise independent of the particular model generating the signal process $y(\cdot)$.

Note: This result agrees with the stationary theory of [1].

No such result holds for smoothing of the state variable, as we shall now see. The Wiener–Hopf equation applicable is

$$E[x(t)z'(\tau)] = \int_{t_0}^{t+\Delta} A_x(t + \Delta,\sigma)E[z(\sigma)z'(\tau)]\, d\sigma, \qquad t_0 \leq \tau < t + \Delta. \quad (54)$$

Let us examine $E[x(t)z'(\tau)]$, in particular for $\tau > t$. We have

$$E[x(t)z'(\tau)] = E[x(t)y'(\tau)] = E[x(t)x'(\tau)]H(\tau).$$

Now let $P(t) = E[x(t)x'(t)]$. As we know, $P(t)$ depends critically on the particular model of the signal process $y(\tau)$. It is not hard to check that for $\tau > t$,

$$E[x(t)x'(\tau)] = P(t)\Phi'(\tau,t)$$

so that

$$E[x(t)z'(\tau)] = [H'(\tau)\Phi(\tau,t)P(t)]'. \quad (55)$$

In order to compute $E[x(t)z'(\tau)]$, knowing $\Phi(\cdot,\cdot)$ and $H(\cdot)$, it is evident from (55) that $P(t)$ itself, rather than, say, $P(t)H(t)$ or $P(t)H(t) + G(t)J_1'(t)$, must be known. Equivalently, we must know the detailed signal process model. Observe, though, that

$$E[y(t)z'(\tau)] = H'(t)E[x(t)z'(\tau)] = H'(t)P(t)\Phi'(\tau,t)H(\tau) \quad (56)$$

from (55). This quantity, appearing in the Wiener–Hopf equation (51), depends on $\Phi(\cdot,\cdot)$, $H(\cdot)$, and PH as distinct from $P(\cdot)$. Reference to Section II shows that

$$E[y(t)y'(\tau)] = H'(t)\Phi(t,\tau)P(\tau)H(\tau)1(t - \tau) + H'(t)P(t)\Phi'(\tau,t)H(\tau)1(\tau - t) \quad (57)$$

so that, given $\Phi(\cdot)$ and $H(\cdot)$, knowledge of $E[y(t)y'(\tau)]$ is equivalent to knowledge of PH; therefore (56) is in harmony with Theorem 6.

Kailath has pointed out to us that a formula in [11] provides an immediate computational procedure for obtaining $y_e(t \mid t + \Delta)$, an estimate of $y(t)$ using $z(\tau)$, for $t_0 \leq \tau \leq t + \Delta$, in terms of the quantities $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$, and other quantities computable from them. The formula is

$$y_e(t \mid t + \Delta) = y_e(t) + \int_t^{t+\Delta} R_{\tilde{y}}(t,\tau)[z(\tau) - y_e(\tau)]\, d\tau \quad (58)$$

where $y_e(t)$ is the usual filtered estimate, and $R_{\tilde{y}}(t,\tau) = E\{[y_e(t) - y(t)][y_e(\tau) - y(\tau)]'\}$, which is readily found for $\tau > t$ to be $R_{\tilde{y}}(t,t)\Phi_g'(\tau,t)$, where $\Phi_g(\cdot,\cdot)$ is the transition matrix associated with the optimal filter. As we have seen, $y_e(t)$, $R_{\tilde{y}}(t,t)$, and $\Phi_g'(\tau,t)$ are all computable from the quantities $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$. The error variance is likewise computable from these quantities since, as shown in [11],

$$E\{[y(t) - y_e(t \mid t + \Delta)][y(t) - y_e(t \mid t + \Delta)]'\} = E\{[y(t) - y_e(t)][y(t) - y_e(t)]'\} - \int_t^T R_{\tilde{y}}(t,\tau)R^{-1}(\tau)R_{\tilde{y}}(\tau,t)\, d\tau. \quad (59)$$

Note that it is also possible to express the quantities on the left sides of (58) and (59) as the solution of readily derivable differential equations.

VIII. Summary and Conclusions

The purpose of this section is to briefly highlight the major results. For filtering of a signal process $y(t)$, all that is required under the key assumption $E[y(t)n'(\tau)] = 0$, for $\tau > t$, is knowledge of $E[y(t)z'(\tau)]$, $t \geq \tau$, and $E[n(t)n'(\tau)]$. When the state-space coordinate basis is defined, these quantities fully determine the optimal filter estimating the state. Moreover, the optimal filter can be calculated assuming knowledge only of these quantities. Filter performance, however, depends in more detail on the model. The error variance in estimating $y(t)$ depends on $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$, while the error variance in estimating the state of the signal process model depends on the particular model.

For smoothing of a signal process $y(t)$, again under the assumption $E[y(t)n'(\tau)] = 0$, for $\tau > t$, knowledge is required of $E[y(t)y'(\tau)]$, $E[y(t)n'(\tau)]$, and $E[n(t)n'(\tau)]$ to compute the optimal smoother and its error variance, and for smoothing of the state of a signal process model, detailed knowledge of the model is required, both to compute the optimal smoother and its error variance.

The question of what data is required to compute optimal predictors and to compute their performance, has not been covered in this paper. However, the difference between prediction and filtering is so slight that the interested reader will have no problem in making the necessary adjustments to the filtering results. In particular, for both state and signal process estimation, the same quantities are required for computing optimal predictors and their performance as are required for computing optimal filters and their performance.

Acknowledgment

The first author would like to thank Prof. A. P. Sage for hospitality extended.

References

[1] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* Cambridge, Mass.: M.I.T. Press, 1949.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, J. Basic Eng.*, ser. D, vol. 83, Mar. 1961, pp. 95–108.

[3] B. D. O. Anderson, J. B. Moore, and S. G. Loo, "Spectral factorization of time-varying covariance functions," *IEEE Trans. Inform. Theory*, vol. IT-15, Sept. 1969, pp. 550–557.

[4] J. B. Moore and B. D. O. Anderson, "Spectral factorization of time-varying covariance functions: the singular case," *Math. Syst. Theory*, vol. 4, 1970, pp. 10–23.

[5] B. D. O. Anderson and J. B. Moore, "State estimation via the whitening filter," in *1968 Joint Automatic Control Conf., Preprints*, pp. 123–129.

[6] ——, "Solution of a time-varying Wiener filtering problem," *Electron. Lett.*, vol. 3, Dec. 1967, pp. 562–563.

[7] R. E. Kalman, "New methods and results in linear filtering and prediction theory," in *Proc. Symp. on Engineering Applications of Probability and Random Functions.* New York: Wiley, 1961.

[8] F. Smithies, *Integral Equations.* New York: Cambridge, 1958.

[9] B. D. O. Anderson and T. Kailath, "The choice of signal-process models in Kalman–Bucy filtering," *J. Math. Anal. Appl.*, to be published.

[10] T. Kailath, "Fredholm resolvents, Wiener–Hopf equations, and Riccati differential equations," *IEEE Trans. Inform. Theory*, vol. IT-15, Nov. 1969, pp. 665–672.

[11] T. Kailath and P. Frost, "An innovations approach to least-squares estimation part II: linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, Dec. 1968, pp. 655–660.