SIAM J. CONTROL
Vol. 12, No. 4, November 1974

OPTIMAL FILTERING OF CONTINUOUS-TIME STATIONARY PROCESSES BY MEANS OF THE BACKWARD INNOVATION PROCESS*

ANDERS LINDQUIST†

Abstract. A new approach to linear least squares estimation of continuous-time (wide sense) stationary stochastic processes is presented. The basic idea is that the relevant estimates can be expressed not only in terms of the usual (forward) innovation process but also in terms of a backward innovation process. The functions determining the optimal filter as well as the error covariance functions are seen to satisfy some differential equations. As an important example the Kalman–Bucy filter is considered. It is demonstrated that the optimal gain matrix can be determined from $2mn$ equations (where n is the dimension of the system and m of the output) rather than $\frac{1}{2}n(n + 1)$ as in the conventional theory. This is an advantage when, as is usually the case, $m \ll n$. These equations were first derived by Kailath, who used a different method. Also they are the continuous-time versions of some equations previously obtained (independently of Kailath) by the author.

1. Introduction. In this paper we consider linear least squares filtering of wide sense stationary stochastic vector processes, where the estimation is based on past observations of the process on an increasing but *finite* time interval. Since therefore the filtering estimate will be a nonstationary process, the weighting function of the filter will be a function of two time variables rather than one as in classical Wiener theory, where observations from the infinite past are assumed to be available. This weighting function satisfies a generalized Wiener–Hopf equation for which no general method of solution is known. Since this is also the case when the process to be estimated is nonstationary, it may seem unnecessarily restrictive to assume stationarity. However, it turns out that this assumption will enable us to give simple differential equations for the weighting function and the error covariance function. These equations are completely characterized by the covariance between the estimation error process and the initial value of the estimated process, and therefore, at least in theory, we have reduced the problem to determining this function of one variable.

The usefulness of our results becomes apparent when applying them to Kalman–Bucy filtering [4] of wide sense stationary processes. It is well known that the computation of the "gain matrix" for such a filter requires the solution of an $n \times n$ matrix Riccati differential equation, where n is the dimension of the system. The number n is usually much larger than the dimension m of the observed process. Our approach will yield $2mn$ nonlinear differential equations instead of the $\frac{1}{2}n(n + 1)$ of the Riccati equation and therefore will supply a more effective algorithm for the gain matrix whenever $m \ll n$. The same equations were recently presented by Kailath [3], who derived them directly from the Riccati equation. However, our approach helps to reveal the fact that the property of the error covariance matrix which makes Kailath's method work holds for wide sense stationary stochastic processes in general, and not only for those realized by a finite-dimensional linear stochastic system.

* Received by the editors February 12, 1973.

† Department of Mathematics, University of Florida, Gainesville, Florida. Now at Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506.

In order to obtain our results, in § 3 we define the *backward innovation processes.* This is in the spirit of a previous paper [6] by the author on filtering of discrete-time processes, and we shall find some similarities in the structures of analogous equations, but also some important differences. So, for example, the equations for the "Kalman gain" are somewhat more complicated in the discrete-time case in that an equation for the "backward error covariance" is required. The reason for this, of course, is that in the continuous-time case the innovation processes can be defined (as we have done) to have constant incremental covariances, while in the discrete-time case they are identical to the error processes, for which no such constancy holds. (Hence the discrete-time counterpart of (4.7) also contains the backward error covariance matrix.)

2. Preliminaries. Let $z(t)$ be the m-dimensional stochastic process

$$z(t) = \int_0^t y(\tau)\,d\tau + w(t) \tag{2.1}$$

defined on $[0, T]$, where $y(t)$ is a zero mean vector process[1] such that

$$\int_0^T E|y(t)|^2\,dt < \infty,$$

and $w(t)$ is a process with zero mean and covariance function

$$E\{w(s)w(t)'\} = I \min(s, t) \tag{2.2}$$

(prime denotes transpose), which implies that $w(t)$ has *orthogonal increments.*

If H is the Hilbert space of all second order stochastic variables (of course, we have tacitly assumed an underlying probability space (Ω, B, P)) with inner product $(\xi, \eta) = E\{\xi\eta\}$, then define $H_t(z)$ to be the closed linear hull in H of the stochastic variables $\{z_i(s);\ 0 \leqq s \leqq t,\ i = 1, 2, \cdots, m\}$. Furthermore, for any $\xi \in H$, let $\hat{E}_t^z\xi$ denote the projection of ξ onto $H_t(z)$, i.e., the wide sense conditional mean of ξ given $\{z(s);\ 0 \leqq s \leqq t\}$. If x is a vector with components $x_i \in H$, we shall take $\hat{E}_t^z x$ to mean the vector with components $\hat{E}_t^z x_i$.

We shall need a few results from linear filtering theory which in the present form are essentially due to Kailath. Denote $\hat{E}_t^z y(t)$ by $\hat{y}(t)$ and define the innovation process

$$v(t) = z(t) - \int_0^t \hat{y}(\tau)\,d\tau, \tag{2.3}$$

for which we have the following lemmas.

LEMMA 2.1. *The process $v(t)$ has zero mean and covariance function* (2.2) *and hence orthogonal increments. Moreover,*

$$H_t(v) = H_t(z), \qquad 0 \leqq t \leqq T. \tag{2.4}$$

LEMMA 2.2. *Let x be a stochastic vector with components in H, and let $v(t)$ be a zero mean vector process with orthogonal increments and covariance function*

[1] In order to take full advantage of integration theory, we assume that all stochastic processes defined are measurable in (t, ω).

(2.2). *Then*

$$\hat{E}_t^v x = \int_0^t \frac{d}{ds} E\{xv(s)'\}\, dv(s). \tag{2.5}$$

Informal versions of these lemmas have appeared in a series of papers by Kailath on the "innovation method". For rigorous proofs see Kailath [2] or Lindquist [5].

3. The backward innovation processes. For the moment assuming that $t \in [0, T]$ is fixed, define the following stochastic processes for $s \in [0, t]$:

$$y_t(s) = y(t - s), \tag{3.1}$$

$$z_t(s) = z(t) - z(t - s), \tag{3.2}$$

$$w_t(s) = w(t) - w(t - s), \tag{3.3}$$

where y, x and w are the processes defined in § 2. Then equation (2.1) gives us

$$z_t(s) = \int_0^s y_t(\tau)\, d\tau + w_t(s), \tag{3.4}$$

which is an equation of the same type as (2.1), for it is immediately clear that

$$E\{w_t(s)w_t(\tau)'\} = I \min(s, \tau). \tag{3.5}$$

Therefore $w_t(s)$ has orthogonal increments for each fixed t. Also it is clear that

$$z(s) = z_t(t) - z_t(t - s) \tag{3.6}$$

and therefore

$$H_t(z_t) = H_t(z). \tag{3.7}$$

Now, introducing the notation $\hat{y}_t(s) = \hat{E}_s^{z_t} y_t(s)$, we consider the innovation process corresponding to (3.4):

$$v_t(s) = z_t(s) - \int_0^s \hat{y}_t(\tau)\, d\tau, \tag{3.8}$$

which we shall call the *backward innovation process* for $\{z(s); 0 \leqq s \leqq t\}$. Clearly we have one such process for each $t \in [0, T]$. The following lemma is then an immediate consequence of Lemma 2.1 and equation (3.7).

LEMMA 3.1. *For each fixed t, $v_t(s)$ has zero mean and covariance function* (3.5), *and hence orthogonal increments. Moreover,*

$$H_s(v_t) = H_s(z_t), \qquad 0 \leqq s \leqq t, \tag{3.9}$$

and, in particular,

$$H_t(v_t) = H_t(z_t) = H_t(z) = H_t(v). \tag{3.10}$$

Therefore, whenever we wish to determine a linear least squares estimate based on the data $\{z(s); 0 \leqq s \leqq t\}$, we can also express it in terms of z_t, v or v_t, whichever we find appropriate.

4. An equation for the error covariance. Let $z(t)$ be the m-dimensional data process defined by (2.1) and $x(t)$ a related n-dimensional zero mean stochastic process such that the compound process $(x(t), y(t))$ is *wide sense stationary* with $E\{x(t)x(t)'\} = P_0$. Also, to simplify matters, assume that (x, y) and w are uncorrelated.

Now, our problem is to determine the linear least squares estimate $\hat{x}(t) = \hat{E}_t^z x(t)$, and in the process of doing so we are interested in the estimation error covariance function

$$P(t) = E\{\tilde{x}(t)\tilde{x}(t)'\}, \tag{4.1}$$

where $\tilde{x}(t) = x(t) - \hat{x}(t)$.

To this end, we recall the well-known fact that (for a fixed t) the filtering estimate $\hat{y}_t(s)$ of $y_t(s)$ given the data $\{z_t(\tau); 0 \leqq \tau \leqq s\}$ is

$$\hat{y}_t(s) = \int_0^s G(s, \tau)\, dz_t(\tau), \tag{4.2}$$

where G is a function (defined through a generalized Wiener–Hopf equation) only of C_t, where $C_t(\tau, s) = E\{y_t(\tau)y_t(s)'\}$. However, due to the stationarity, $C_t(\tau, s) = E\{y(s)y(\tau)'\}$ does not depend on the parameter t, and hence G is good for all $t \in [0, T]$. Therefore we have

$$\begin{aligned} E\{x(t)\hat{y}_t(s)'\} &= E\left\{x(t)\left[\int_0^s G(s, \tau)y(t - \tau)\, d\tau\right]'\right\} \\ &= E\left\{x(s)\left[\int_0^s G(s, \tau)y(s - \tau)\, d\tau\right]'\right\} \\ &= E\{x(s)\hat{y}_s(s)'\}, \end{aligned} \tag{4.3}$$

where we have used the fact that (x, y) is wide sense stationary and x and w are uncorrelated.

We are now in a position to apply Lemmas 2.2 and 3.1 to see that

$$\hat{x}(t) = \int_0^t Q(t, s)\, dv_t(s), \tag{4.4}$$

where

$$\begin{aligned} Q(t, s) &= E\{x(t)[y_t(s) - \hat{y}_t(s)]'\} \\ &= E\{x(s)[y_s(s) - \hat{y}_s(s)]'\} \\ &= E\{\tilde{x}(s)y(0)'\}, \end{aligned} \tag{4.5}$$

where again we have exploited the stationarity and uncorrelatedness properties mentioned above, relation (4.3), and also the fact that $\hat{x}(s)$ and $y_s(s) - \hat{y}_s(s)$ are orthogonal and that the same is true for $\tilde{x}(s)$ and $\hat{y}_s(s)$. Hence $Q(t, s)$ does not depend on t, and we shall therefore call it $Q(s)$:

$$\hat{x}(t) = \int_0^t Q(s)\, dv_t(s). \tag{4.6}$$

Then we have the error covariance

$$
\begin{aligned}
P(t) &= E\{x(t)x(t)'\} - E\{\hat{x}(t)\hat{x}(t)'\} \\
&= P_0 - \int_0^t Q(s)Q(s)'\,ds,
\end{aligned}
\tag{4.7}
$$

which concludes the proof of the following theorem.

THEOREM 4.1. *With the conditions imposed in the beginning of this section, the error covariance* (4.1) *satisfies the matrix differential equation*

$$
\begin{aligned}
\dot{P}(t) &= -Q(t)Q(t)', \\
P(0) &= P_0,
\end{aligned}
\tag{4.8}
$$

where

$$
Q(t) = E\{\tilde{x}(t)y(0)'\}. \tag{4.9}
$$

Here the dynamics of the $n \times m$ matrix function Q is essentially that of the error signal $\tilde{x}$, so to proceed we have to impose further conditions on the process $x(t)$. We choose to illustrate this by applying Theorem 4.1 to the Kalman–Bucy filter.

Assume that the n-dimensional wide sense stationary process $x(t)$ is given by the stochastic differential equation

$$
dx = Ax\,dt + B\,dv, \qquad x(0) = x_0, \tag{4.10}
$$

and the m-dimensional data process by

$$
dz = Hx\,dt + dw, \qquad z(0) = 0,
$$

so that $y(t)$ is in fact equal to $Hx(t)$. Here $v(t)$ is a vector process of type (2.2), x_0 is a zero mean stochastic variable, and x_0, v and w are pairwise uncorrelated. The matrices A, B and H are constant.

Now, it is well known [4] that the filtering estimate $\hat{x}(t)$ is generated by

$$
d\hat{x} = A\hat{x}\,dt + K(dz - H\hat{x}\,dt), \qquad \hat{x}(0) = 0, \tag{4.11}
$$

where the "gain-matrix" function K is given by

$$
K(t) = P(t)H'. \tag{4.12}
$$

The $n \times n$ matrix function P is usually determined from a matrix Riccati equation, which amounts to solving $\frac{1}{2}n(n + 1)$ nonlinear differential equations, in order to obtain the nm functions in the gain matrix K. Our procedure yields $2mn$ equations, which is a major advantage whenever, as is often the case, $m \ll n$. To see this, first observe that the error process $\tilde{x}(t)$ is given by

$$
d\tilde{x} = (A - KH)\tilde{x}\,dt + B\,dv - dw \tag{4.13}
$$

with initial condition $\tilde{x}(0) = x_0$. Moreover, $y(0) = Hx_0$, which is uncorrelated with v and w. It is then easy to see that

$$
\dot{Q} = (A - KH)Q, \qquad Q(0) = P_0H',
$$

and that we therefore have the following $2mn$ equations to determine the optimal gain:

$$\begin{aligned} \dot{K}(t) &= -Q(t)Q(t)'H', \\ \dot{Q}(t) &= (A - K(t)H)Q(t), \end{aligned} \tag{4.14}$$

with initial conditions $K(0) = Q(0) = P_0H'$. Clearly, there exists a *unique* solution of the system (4.14). Indeed, establishing this is a standard exercise in the use of the contraction mapping principle.

Equations (4.14) have also been obtained by Kailath [3] by differentiating the Riccati equation. However, unlike Kailath's method, ours completely avoids the Riccati equation. It also demonstrates the fact that the low rank property (in the interesting case $m \ll n$) of $\dot{P}$ is not only a property of "lumped" stationary processes (4.10) but is one of stationary processes in general.

5. Differential equations for the weighting function. Consider the problem to determine the linear least squares estimate $\hat{y}(t) = \hat{E}_t^z y(t)$, where as before the data process $z(t)$ is given by (2.1):

$$z(t) = \int_0^t y(\tau)\,d\tau + w(t). \tag{5.1}$$

We assume that y is wide sense stationary and that y and w are uncorrelated. Then as we pointed out in § 4, we have

$$\hat{y}(t) = \int_0^t F(t, t - s)\,dz(s), \tag{5.2}$$

where F is the weighting function to be determined. This function is known to satisfy a Fredholm integral equation (a generalized Wiener–Hopf equation), but we shall demonstrate that it also satisfies a system of differential equations.

Now it is easily seen that we can rewrite (5.2) in terms of the backward data process (3.2) to obtain

$$\hat{y}(t) = \int_0^t F(t, s)\,dz_t(s). \tag{5.3}$$

Also we define the backward weighting function F^* by the equation

$$\hat{y}_t(s) = \int_0^s F^*(s, s - \tau)\,dz_t(\tau). \tag{5.4}$$

As we pointed out in § 4, the stationarity insures that F^* is the same for all values of the parameter t. Moreover,

$$\hat{y}_t(t) = \int_0^t F^*(t, \tau)\,dz(\tau). \tag{5.5}$$

We can also express $\hat{y}$ and $\hat{y}_t$ in terms of the innovation processes v and v_t.

In fact, equation (4.6) yields for the case $x = y$:

$$\hat{y}(t) = \int_0^t \Gamma(s)\, dv_t(s), \tag{5.6}$$

where

$$\Gamma(t) = E\{\tilde{y}(t)y(0)'\} \tag{5.7}$$

or, introducing the notation $\tilde{y}_t(s) = y_t(s) - \hat{y}_t(s)$,

$$\Gamma(t) = E\{\tilde{y}(t)\tilde{y}_t(t)'\}. \tag{5.8}$$

Also, a straightforward application of Lemma 2.2 (with $v = \nu$) yields

$$\hat{y}_t(t) = \int_0^t \Gamma(s)'\, d\nu(s). \tag{5.9}$$

Hence, by (3.8),

$$\begin{aligned}\hat{y}(t) &= \int_0^t \Gamma(s)\left[dz_t(s) - \int_0^s F^*(s, s-\tau)\, dz_t(\tau)\, ds\right]\\ &= \int_0^t \left[\Gamma(s) - \int_s^t \Gamma(\tau)F^*(\tau, \tau - s)\, d\tau\right] dz_t(s)\end{aligned}$$

and, by (2.3),

$$\begin{aligned}\hat{y}_t(t) &= \int_0^t \Gamma(s)'\left[dz(s) - \int_0^s F(s, s-\tau)\, dz(\tau)\, ds\right]\\ &= \int_0^t \left[\Gamma(s)' - \int_s^t \Gamma(\tau)'F(\tau, \tau - s)\, d\tau\right] dz(s).\end{aligned}$$

(The change of the order of integration is permitted due to a Fubini-type theorem for stochastic integrals. See, e.g., [1, p. 197].) It is clear from these expressions that

$$F(t, s) = \Gamma(s) - \int_s^t \Gamma(\tau)F^*(\tau, \tau - s)\, d\tau, \tag{5.10}$$

$$F^*(t, s) = \Gamma(s)' - \int_s^t \Gamma(\tau)'F(\tau, \tau - s)\, d\tau. \tag{5.11}$$

Now, the following theorem is a continuous-time analogue of Lemmas 3.1 and 3.2 in [6].

THEOREM 5.1. *The weighting functions F and F^* satisfy the following differential equations*:

$$\frac{\partial F}{\partial t}(t, s) = -\Gamma(t)F^*(t, t - s), \tag{5.12}$$

$$\frac{\partial F^*}{\partial t}(t, s) = -\Gamma(t)'F(t, t - s) \tag{5.13}$$

for $t \geqq s$, with initial conditions $F(s,s) = \Gamma(s)$ and $F^*(s,s) = \Gamma(s)'$, where Γ is defined by (5.7) *or* (5.8). *The error covariances $R(t) = E\{\tilde{y}(t)\tilde{y}(t)'\}$ and $R^*(t) = E\ \{\tilde{y}_t(t)\tilde{y}_t(t)'\}$ satisfy*:

$$\dot{R}(t) = -\Gamma(t)\Gamma(t)', \tag{5.14}$$

$$\dot{R}^*(t) = -\Gamma(t)'\Gamma(t), \tag{5.15}$$

with initial conditions $R(0) = R^*(0) = E\{y(t)y(t)'\}$.

Moreover, if x is the process of Theorem 4.1 *and N is the weighting function defined by*

$$\hat{x}(t) = \int_0^t N(t, t-s)\,dz(s), \tag{5.16}$$

then N satisfies

$$\frac{\partial N}{\partial t}(t,s) = -Q(t)F^*(t, t-s) \tag{5.17}$$

for $t \geqq s$, with initial condition $N(s,s) = Q(s)$, where Q is given by (4.9).

Proof. Equations (5.12) and (5.13) follow from (5.10) and (5.11), and (5.17) is derived in the same way as (5.12) only exchanging (5.6) for (4.6). Finally, (5.14) and (5.15) are consequences of Theorem 4.1. To obtain (5.14), put $x = y$, and to obtain (5.15), put $x = y_t$ and exchange y for y_t.

As an example, we can now apply equations (5.12) and (5.13) to obtain an alternative derivation of equations (4.14). In fact, by (5.2) and (5.5), $K(t) = E\{x(t)\tilde{y}(t)'\}$ and $Q(t) = E\{x(t)\tilde{y}_t(t)'\}$ can be expressed in terms of F and F^* respectively. Also observe that $E\{x(t)y(s)'\} = e^{A(t-s)}P_0H'$ and finally that $\Gamma(t) = HQ(t)$.

REFERENCES

[1] I. I. Gikman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*, Saunders, London, 1969.

[2] T. Kailath, *A note on least squares estimation by the innovations method*, this Journal, 10 (1972), pp. 477–486.

[3] ———, *Some Chandrasekhar-type algorithms for quadratic regulators*, Proc. IEEE Conference on Decision and Control, New Orleans, December 1972.

[4] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Trans. ASME Ser. D J. Basic Engrg., 83 (1961), pp. 95–108.

[5] A. Lindquist, *Optimal control of linear stochastic systems with applications to time lag systems*. Information Sci., 5 (1973), pp. 81–126.

[6] ———, *A new algorithm for optimal filtering of discrete-time stationary processes*, this Journal, 12 (1974), pp. 736–746.