THE OPTIMAL FILTERING OF MARKOV JUMP
PROCESSES IN ADDITIVE WHITE NOISE

by

M. Zakai

15 June 1965

Research Note No. 563

Applied Research Laboratory
Sylvania Electronic System
A Division of Sylvania Electric Products Inc.
Waltham, Massachusetts 02154
The optimal filtering of Markov jump processes in additive white noise.

This note is based on Wonham [1]. The differences between this note and [1] are discussed in Section VIII.

I. Statement of the Problem.

Let \( x(t) \) be Markov jump process with stationary transition probabilities and with a finite number of states. Let \( a_1, a_2, \ldots, a_K \) be the states, let \( p_{ij}(h) \) be the transition probabilities

\[
p_{ij}(h) = \text{Prob}\left\{ x(t+h) = a_j \bigg| x(t) = a_i \right\}.
\]

Let

\[
p_{ij}(h) = \begin{cases} 
1 - \nu_i h + o(h) & i=j, \ h \to 0 \\
\nu_{ij}(h) + o(h) & i \neq j, \ h \to 0
\end{cases}
\]

where \( \nu_i > 0, \ \nu_{ij} \geq 0 \) and \( \nu_i = \sum_{j=1}^{K} \nu_{ij}, \ i = 1, \ldots, K. \)

Let \( p_i(0) \) be the initial distribution of \( x(0) \). In addition let \( y(t) \) be a process given by

\[
dy(t) = x(t) \, dt + \beta \, dw(t), \quad y(0) = 0
\]

where \( \beta \) is known and \( w(t) \) is a standard Brownian motion. The problem is to find

\[
p_j(t) = \text{Prob}\left\{ x(t) = a_j \bigg| \pi^t_{0,y}(\cdot) \right\}
\]

where \( \pi^t_{a,y}(\cdot) \) stands for \( y(s), \ a \leq s \leq b. \)
II. A general expression for \( p_j(t) \)

The conditional probability \( \text{Prob}\{ x(t) = a_j \mid A \} \) (where \( A \) is some condition) is the same as the conditional expectation of the function \( \delta_{ij} \) (\( j \) fixed) given \( A \). Applying Doob’s theorem 8.8 ([2, p. 21]), there exists a sequence \( t_1, t_2, \ldots \), all in \([0, t]\), such that a.s.

\[
p_j(t) = \text{Prob}\{ x(t) = a_j \mid y(t_1), y(t_2), \ldots \}
\]

and by the martingale convergence theorem ([2, Cor 1 p. 332]) a.s.

\[
p_j(t) = \lim_{n \to \infty} \theta^{(n)}_j(t) \tag{1}
\]

where \( \theta^{(n)}_j(t) = \text{Prob}\{ x(t) = a_j \mid y(t_1), y(t_2), \ldots, y(t_n) \} \).

In the following we will use:

\[
\text{Prob}\{ x(t) \in A \mid \pi_0^0 y(\cdot) \in B \} = \frac{\text{Prob}\{ x(t) \in A, \pi_0^0 y(\cdot) \in B \}}{\text{Prob}\{ \pi_0^0 y(\cdot) \in B \}}.
\]

Consider now a fixed \( t_\mu, 0 < t_\mu < t \); let \( p_j(t \mid y(t_\mu)) \) be the probability that \( x(t) = a_j \) given \( y(t_\mu) \) and let \( \xi_\mu \) be

\[
\xi_\mu = \int_0^{t_\mu} x(t) \, dt.
\]

Then

\[
p_j \left( t \mid y(t_\mu) \right) = \frac{\sum_{i=1}^K p_i(0) p_{ij}(t) \int_{-\infty}^\infty P_1(y(t_\mu) - \xi_\mu) P_2(\xi_\mu \mid x(0) = a_i, x(t) = a_j) \, d\xi_\mu}{\sum_{j=1}^K (\text{The same expression as in the numerator})}.
\]

Note that \( P_1 \) is normal \((0, \beta^2 t_\mu)\). Let \( \tilde{x}(t) \) be a process independent of \( x(t) \) and \( w(t) \) with the same law as \( x(t) \). Let \( \tilde{\xi}(s) = \int_0^s \tilde{x}(t) \, dt \), then
\[ p_j \left( t \mid y(t) \right) = \frac{\sum_{i=1}^{K} p_i(0) p_{ij}(t) E \left\{ \exp \left[ -\frac{(y(t) - \bar{\xi}(t))^2}{2\beta^2 t} \right] \bigg| \bar{x}(0) = a_i, \bar{x}(t) = a_j \right\}}{\sum_{j=1}^{K} \sum_{i=1}^{K} p_i(0) p_{ij}(t) E \left\{ \exp \left[ -\frac{(y(t) - \bar{\xi}(t))^2}{2\beta^2 t} \right] \bigg| \bar{x}(0) = a_i, \bar{x}(t) = a_j \right\}}. \]

The conditioning of the expectations in the above expression are all \( \bar{x}(s) \) paths which start at \( s = 0 \) with \( a_i \) and terminate at \( s = t \) in the state \( a_j \). Now let \( s_{r,n} = r \frac{t}{n}, \ r = 0, 1, \ldots, n \).

Let
\[ \eta_{r,n} = y(s_{r+1,n}) - y(s_{r,n}), \]
\[ \tilde{\xi}_{r,n} = \int_{s_{r,n}}^{s_{r+1,n}} \bar{x}(s) \, ds. \]

Then by the same arguments as above
\[ p_j \left( t \mid y(r \frac{t}{n}), r = 0, 1, \ldots, n \right) = \]
\[ = \frac{\sum_{i=1}^{K} p_i(0) p_{ij}(t) E \left\{ \exp \left[ \sum_{r=0}^{n-1} \frac{(\eta_{r,n} - \tilde{\xi}_{r,n})^2}{2\beta^2 t/n} \right] \bigg| \bar{x}(0) = a_i, \bar{x}(t) = a_j \right\}}{\sum_{j=1}^{K} \text{(numerator))}}. \] (2)

The argument of the exponential is
\[ = \sum_{r=0}^{n-1} \left( \eta_{r,n}^2 - 2\eta_{r,n}\tilde{\xi}_{r,n} + \tilde{\xi}_{r,n}^2 \right) \frac{1}{2\beta^2 t/n}. \]

The first term will be cancelled by the same term in the denominator. The last term converges a.s. as \( n \to \infty \) to
\[ \frac{1}{2\beta^2} \int_0^t \bar{x}(t)^2 \, dt. \]

The middle term converges a.s. to
\[ -\frac{1}{\beta^2} \int_0^t \bar{x}(t) \, dy(t). \]
We want to apply these results to the evaluation of the limit of the numerator of (2) as
\( n \to \infty \). In order to do that we have to show that if \( f_n \to f \) as \( n \to \infty \), then \( E(f_n | ) \to E(f | ) \).

Since
\[
\exp \left[ - \sum_{r=0}^{n-1} \left( -2\eta_{r,n} \tilde{\xi}_{r,n} + \tilde{\xi}_{r,n}^2 \right) \right] \leq \exp \left[ A_{\text{max}}^2 t + 2A_{\text{max}} \cdot \sup \{ |y(t_1) - y(t_2)|, t_1, t_2 \in [0, t] \} \right]
\]
it follows, by dominated convergence that the limit of the numerator of (2), as \( n \to \infty \), is
\[
\psi_j(t) = \sum_{i=1}^{K} p_i(0) p_{ij}(t) E \left\{ \exp \left[ - \frac{1}{2\beta^2} \int_0^t \tilde{x}^2(s) \, ds + \frac{1}{\beta^2} \int_0^t \tilde{x}(s) \, dy(s) \right] \mid \tilde{x}(0) = a_i, \tilde{x}(t) = a_j \right\}
\]
and the conditioning is with respect to all the paths which start at \( \tilde{x}(0) = a_i \) and terminate at \( \tilde{x}(t) = a_j \). Similarly the limit of the denominator is \( \sum_{i=1}^{K} \psi_i(t) \) where \( \psi_j(t) \) is given by equation (3). Since \( \psi_j(t) > 0 \) a.s. we have
\[
\lim_{n \to \infty} p_j(t \mid y(r_{\frac{t}{n}}), r = 0, 1, \ldots, n) = \frac{\psi_j(t)}{\sum_{i=1}^{K} \psi_i(t)}.
\]
(4)

The limits (3) and (4) were obtained by a particular sequence of partitions of \([0, t]\), but it is clear that the same result will hold for any sequence of partitions \( \{s_r,n\} \) such that
\[
0 = s_{0,n} < s_{1,n} < \ldots < s_{n,n} = t \text{ and such that } \max_r (s_{r+1,n} - s_{r,n}) \to 0 \text{ as } n \to \infty.
\]
We may therefore use a sequence for which (4) is true. Therefore
\[
p_j(t) = \frac{\psi_j(t)}{\sum_{i=1}^{K} \psi_i(t)}.
\]
(5)
III. The stochastic differential equation for $\psi_j(t)$.

Let $\psi_{i,j}(a, b), \ (b > a)$ be

$$\psi_{i,j}(a, b) = p_{ij}(b-a) \left\{ \exp \left[ -\frac{1}{2\beta^2} \int_a^b \bar{x}^2(s) \, ds + \frac{1}{\beta^2} \int_a^b \bar{x}(s) \, dy(s) \right] \mid \bar{x}(a) = a_i, \bar{x}(b) = a_j \right\}.$$  

Then, comparing with (3):

$$\psi_j(t) = K \sum_{i=1}^K p_{i}(0) \psi_{i,j}(0, t).$$

Consider a fixed realization of $\pi^{t+h}_0 y(\cdot)$, we prove now that

$$\psi_j(t+h) = \sum_{i=1}^K \psi_i(t) \cdot \psi_{i,j}(t, t+h). \quad (6)$$

Proof:

$$\psi_{i,j}(0, t+h)$$

$$= p_{ij}(t+h) E \left\{ \exp \left[ -\int_0^{t+h} \ldots + \int_0^{t+h} \ldots \right] \mid \bar{x}(t+h) = a_j, \bar{x}(0) = a_i \right\}$$

$$= p_{ij}(t+h) \sum_{k=1}^K E \left\{ \exp \left[ -\int_0^{t+h} \ldots + \int_0^{t+h} \ldots \right] \mid \bar{x}(t+h) = a_j, \bar{x}(t) = a_k, \bar{x}(0) = a_i \right\}$$

$$\cdot \text{Prob} \left\{ \bar{x}(t) = a_k \mid \bar{x}(t+h) = a_j, \bar{x}(0) = a_i \right\}. $$

Since $\bar{x}(t)$ is a Markov process, the conditional expectation becomes the product of two conditional expectations (since, given $\bar{x}(t), \bar{x}(t-\alpha)$ and $\bar{x}(t+\beta)$ are independent for $\alpha, \beta > 0$).

Moreover

$$\text{Prob} \left\{ \bar{x}(t) = a_k \mid \bar{x}(t+h) = a_j, \bar{x}(0) = a_i \right\} = \frac{p_{ik}(t) p_{kj}(h)}{p_{ij}(t+h)}.$$
Therefore

\[ \psi_{i,j}(0, t+h) = \sum_{k=1}^{K} p_{ik}(t) E \left\{ \exp \left[ -\int_{0}^{t} \ldots + \int_{0}^{t} \mid \bar{x}(t) = a_k, \bar{x}(0) = a_i \right] \right\} \]

\[ \times \quad p_{kj}(h) E \left\{ \exp \left[ -\int_{0}^{t} \ldots + \int_{0}^{t} \mid \bar{x}(t+h) = a_j, \bar{x}(t) = a_k \right] \right\} \]

\[ = \sum_{k=1}^{K} \psi_{i,k}(0, t) \cdot \psi_{k,j}(t, t+h) \]

which is the required result.

Since the \( x(t) \) process is Markov and the \( w(t) \) process has independent increments, it follows that \( \psi_{i,k}(0, t) \) and \( \psi_{k,j}(t, t+h) \) are conditionally independent given \( x(t) \). Therefore the process \( (x(t), \psi_1(t), \psi_2(t), \ldots, \psi_K(t)) \) is a \( K+1 \) dimensional Markov process.

\( \psi_{i,j}(t, t+h) \) will now be evaluated for small \( h \). Assuming that \( h \) is small enough so that the possibility that more than one transition in \([t, t+h]\) can be ignored we have

\[ \psi_{i,i}(t, t+h) \cong (1 - \nu_i h) \exp \left[ -\frac{a_i^2}{2\beta^2} h + \frac{a_i}{\beta^2} (y(t+h) - y(t)) \right] \]

\[ \psi_{i,j}(t, t+h) \cong \nu_{ij} h \exp \left[ -\frac{a_j^2}{2\beta^2} h \theta_1 + \frac{a_j}{\beta^2} \theta_2 (y(t+h) - y(t)) \right] \quad j \neq i \]

The factors \( \theta_1 \) and \( \theta_2 \) were included in the last expression in order to indicate that it is unknown where in \([t, t+h]\) the transition occurred; it will turn out that this is immaterial.

Setting now \( y(t+h) - y(t) = \int_{t}^{t+h} x(s) ds + \beta w(t+h) - \beta w(t) \) and expanding the exponential in a power series we obtain

\[ \psi_{i,i}(t, t+h) \cong 1 - \nu_i h - \frac{a_i^2}{2\beta^2} h + \frac{a_i}{\beta^2} (y(t+h) - y(t)) + \frac{a_i^2}{2\beta^4} \beta^2 (w(t+h) - w(t))^2 + o_1(h, (\Delta w)^2) \]

where \( o_1(h, (\Delta w)^2) \) denotes the terms omitted. Also
\[
\psi_{i,j}(t, t+h) \equiv \nu_{ij} h + o_2(h, (\Delta_h w)^2).
\]

Substituting into (6) we have

\[
\psi_j(t+h) = \psi_j(t) - \nu_j \psi_j(t) h - \frac{a_j^2}{2\beta^2} h \psi_j(t) + \frac{a_j \psi_j(t)}{\beta^2} (y(t+h) - y(t)) + \frac{a_j^2}{2\beta^2} (w(t+h) - w(t))^2 + \sum_{i=1}^{K} \psi_i(t) \nu_{ij} h + o_3(h, (\Delta_h w)^2).
\]

Therefore

\[
\psi_j(t) - \psi_j(0) = \int_0^t \left[ -\nu_j \psi_j(s) + \sum_{i=1}^{K} \psi_i(s) \nu_{ij} - \frac{a_j^2 \psi_j(s)}{2\beta^2} \right] ds + \int_0^t \frac{a_j \psi_j(s)}{\beta^2} dy(s) + \lim_{h \to 0} \sum_{r=1}^{t/h} \frac{a_j^2 \psi_j(hr)}{2\beta^2} (w((r+1)h) - w(rh))^2 + \lim_{h \to 0} \sum_{r=1}^{t/h} o_3(h, (\Delta_h w)^2).
\]

(7)

The first sum can be shown \[6\] to converge a.s. to

\[
\int_0^t \frac{a_j^2 \psi_j(s)}{2} ds.
\]

The second sum can be shown \[3\] to converge to 0. (For example, if \(|f(s)| \leq M\) in \([0, t]\) then the sum \(\sum_{r=1}^{t/h} f(hr) h (w((r+1)h) - w(rh))\) is bounded by

\[
Mt \sup_t |w(t+h) - w(t)|
\]

but since \(w(t)\) is a.s. continuous it is also uniformly continuous and the last term converges to zero as \(h \to 0\), hence the sum converges a.s. to zero).

Equation (7) becomes:

\[
d\psi_j(t) = -\nu_j \psi_j(t) dt + \sum_{i=1}^{K} \nu_{ij} \psi_i(t) dt + \frac{a_j \psi_j(t)}{\beta^2} dy(t),
\]

(8)
with the initial condition \( \psi_j(0) = p_j(0) \).

Let \( \phi(t) = \sum_{j=1}^{K} \psi_j(t) \). Then

\[
d\phi(t) = \sum_{i=1}^{K} \frac{a_i \psi_i(t)}{\beta^2} \, dy(t),
\]

and

\[
p_j(t) = \frac{\psi_j(t)}{\sum_{i=1}^{K} \psi_i(t)} = \frac{\psi_j(t)}{\phi(t)}.
\]

Equation (8) is a stochastic differential equation for \( \psi_j(t) \) from which \( p_j(t) \) can be obtained by (9).

The Langevin equation corresponding to (8) can be derived using equation (4.30) of [5]; the result is

\[
\frac{d\psi_j(t)}{dt} = -\nu_j \psi_j(t) + \sum_{i \neq j} \nu_{ij} \psi_i(t) + \frac{1}{2} \frac{a_j^2 \psi_j(t)}{\beta^2} + \frac{a_j \psi_j(t)}{\beta^2} \frac{dy(t)}{dt} + a_j \psi_j(t) \beta^2 dt \tag{10}
\]

and \( dy(t)/dt \) is \( x(t) \) plus “white noise”.

IV. The stochastic differential equation for \( p_j(t) \).

Since (by definition of \( \psi_j \) and \( \phi \)) \( \phi(t) \neq 0 \) a.s., we may apply Ito’s rule of differentiation [3] to (9):

\[
dp_j(t) = \frac{d\psi_j(t)}{\phi(t)} - \frac{\psi_j(t) d\phi(t)}{\phi^2(t)} - \frac{1}{2} \frac{a_j \psi_j(t)}{\beta^2} \frac{d\phi(t)}{\phi(t)} - \frac{a_j \psi_j(t)}{\beta^2} \frac{dy(t)}{dt} \tag{11}
\]

Substituting for \( d\psi_j \) and \( d\phi \), and setting \( \sum_{i=1}^{K} a_i \psi_i(t) \phi(t) = \overline{\pi}(t) \) we get

\[1\text{see [3] or [7]}

\[ dp_j(t) = -\nu_j p_j(t) \, dt + \sum_{i \neq j}^K p_i(t) \nu_{ij} \, dt \]
\[ + \beta^{-2} (a_j - \overline{x}(t)) \, p_j(t) \, dy(t) \]
\[ + \beta^{-2} \overline{x}(t) \, p_j(t) \, (a_j - \overline{x}(t)) \, dt \]
\[ \text{[}\overline{x}(t) = \sum_{j=1}^K a_j p_j(t)\text{]} \]

which are the equations derived by Wonham. Note that as \( \beta^2 \to \infty \) we get the Kolmogorov forward equation (as expected). The equations for \( \psi_j(t) \) are more elegant than those for \( p_j(t) \). However, the equations for \( p_j(t) \) are probably more useful for applications since they ensure that the output is always in \([0, 1]\) (while \( \psi_j(t) \) can be anywhere in \((0, \infty)\)). It seems also that perhaps the \( p_j(t) \) may have a stationary distribution while \( \psi_j(t) \) may not have such a distribution.

The Langevin equation corresponding to (11) is, by eq. (4.30) of [5],
\[
\frac{dp_j(t)}{dt} = -\nu_j p_j + \sum_{i \neq j} p_i \nu_{ij} + \frac{1}{2} p_j \beta^{-2} \left( a_j^2 - \sum_{i=1}^K a_i^2 p_i \right) \\
+ \beta^{-2} (a_j - \overline{x}) \, p_j \, \frac{dy}{dt} .
\]

\[ \text{(12)} \]

V. Example - The random telegraph signal.

In this case \( a_1 = 1, \ a_2 = -1 \)
\( \nu_i = \nu_{ij} = \nu; \ i,j = 1, 2 \)

where \( \nu \) is the expected number of jumps.

Let \( q(t) = p_1(t) - p_2(t) \)

then \( \overline{x}(t) = q(t) \) and the equations of the last paragraph become
\[
dq(t) = -2\nu q(t) \, dt - \beta^{-2} q(t)(1 - q^2(t)) \, dt + \beta^{-2}(1 - q^2(t)) \, dy(t) \\
\]

or, equivalently \( dq(t) = -2\nu q(t) \, dt - \beta^{-2} q(t)(1 - q^2(t)) \, dt + \beta^{-2}(1 - q^2(t)) \, x(t) \, dt \)
\[
+ \beta^{-1}(1 - q^2(t)) \, dw(t) .
\]
The Langevin equivalent of this stochastic differential equation is given by \cite{5, 7}

\[
\frac{dq(t)}{dt} = -2\nu q(t) + \beta^{-2}(1-q^2(t)) x(t) + \beta^{-2}(1-q^2(t)) n(t)
\]

where \(n(t)\) is “white noise”. Let \(x(t) + n(t) = r(t)\), then we have the Riccati equation:

\[
\frac{dq(t)}{dt} = -2\nu q(t) + \beta^{-2}(1-q^2(t)) r(t) \quad [r(t) = x(t) + n(t)].
\]

The physical filter to compute \(q(t)\) will therefore be

\[\begin{array}{c}
\text{multiplier} \\
\text{initial condition } q(0) \\
r(t) \\
\sum \\
\text{amplifier} \\
-2\nu \\
\int \text{dt} \\
\text{integrator} \\
q(t) \\
\text{decision} \\
(1-q^2) \beta^{-2}
\end{array}\]

or:

\[\begin{array}{c}
r(t) \\
\text{linear network with} \\
\text{transfer function} \\
\frac{1}{i\omega + 2\nu} \\
(1-q^2) \beta^{-2} \\
q(t) \\
\text{decision}
\end{array}\]
If, instead of the analog filter we use a digital computer we have to distinguish between two cases. Let \( \omega_n \) be the cutoff frequency of the “white noise” and \( \omega_s \) be the sampling frequency of the computer. Case 1: \( \nu \ll \omega_n \ll \omega_s \), Case 2: \( \nu \ll \omega_s \ll \omega_n \). It follows from \[5\] and \[7\] that in case 1 the computer should be programmed to solve the Langevin equation. In case 2, Maruyama’s approximation theorem is applicable \[5\] and the computer should be programmed to solve Ito’s equation (via Maruyama’s approximation).

An error analysis for this example is discussed in \[1\].

VI. Some transformations on \( \psi_j(t) \).

Equation (10) can be rewritten as

\[
\frac{d\psi(t)}{dt} = A \cdot \psi(t) + \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_K \end{bmatrix} \begin{bmatrix} x(t) + n(t) \\ \beta^2 \end{bmatrix} \cdot \psi(t) \tag{10a}
\]

where \( A \) is a constant matrix and \( \psi \) is the vector \((\psi_1, \ldots, \psi_K)^T\).

Setting

\[ \Gamma(t) = e^{-At} \cdot \psi(t) \]

we get the Langevin equation for \( \Gamma(t) \):

\[
\frac{d\Gamma(t)}{dt} = e^{-At} \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ a_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_K \end{bmatrix} e^{At} \begin{bmatrix} x(t) + n(t) \\ \beta^2 \end{bmatrix} \cdot \Gamma(t). \tag{13}
\]

Setting \( \theta_j(t) = \log \psi_j(t) \)

hence

\[
\frac{d\theta_j(t)}{dt} = -\nu_j + \frac{1}{2} \frac{a_j^2}{\beta^2} + \sum_{i \neq j} \nu_{ij} e^{\theta_i(t)-\theta_j(t)} + \frac{a_j}{\beta^2} (x(t) + n(t)) \tag{14}
\]

and since \( \theta_i \) does not appear in front of the last term, this is the Ito as well as the Langevin equation for \( \theta_j(t) \).
VII. The prediction of $x(t)$.

The problem is now to find the probability that $x(t+h) = a_j$, $h > 0$, conditioned on $\pi_0^t y(\cdot)$. The result follows directly from

$$\text{Prob} \left\{ x(t+h) \in A \mid \pi_0^t y(\cdot) \in B \right\} = \sum_{i=1}^{K} \text{Prob} \left\{ x(t) = a_i \mid \pi_0^t y(\cdot) \in B \right\} \cdot \text{Prob} \left\{ x(t+h) \in A \mid x(t) = a_i, \pi_0^t y(\cdot) \in B \right\} .$$

Since $\text{Prob} \left\{ x(t+h) \mid x(t), \pi_0^t y(\cdot) \right\} = \text{Prob} \left\{ x(t+h) \mid x(t) \right\}$, it follows that (as expected)

$$\text{Prob} \left\{ x(t+h) = a_j \mid \pi_0^t y(\cdot) \right\} = \sum_{i=1}^{K} p_i(t) p_{ij}(h) \quad (15)$$

where $p_i(t)$ is the solution to (11) and $p_{ij}(h)$ are defined at the beginning of section I. The extension of (15) to the probability distribution of functionals on $x(s)$, $s \geq t$, conditioned on $\pi_0^t y(\cdot)$ is obvious.

VIII. Remarks.

Section I, II and the first halves of III and V follow from Wonham [1]. Instead of proceeding directly to obtain the stochastic differential equations for $p_j(t)$ as done in [1] we first derive the stochastic differential equations for $\psi_j(t)$ (section III) from which the stochastic differential equations for $p_j(t)$ are derived by a singular transformation (section IV). The equations for $\psi_j(t)$ are considerably simpler and are of a standard form (section VI). It is believed that a similar approach can be used in the case treated by Stratonovich and Kushner (where $x(t)$ is a diffusion process). The treatment in this note is restricted to $\beta = \text{const}$, the treatment in [1] is for $\beta = \beta(t)$ where $\beta(t)$ is continuously differentiable and bounded away from zero. The extension of the arguments and results of this note to $\beta = \beta(t)$ is straightforward.
A question which was left open in [1] was the problem of the realization of the results as physical “filters”. Recent work related to this problem [5], [7] gives answers to this question. Equation (4.30) of [5], which was used in this note, was derived in [5] by a heuristic argument. Unpublished calculations (for piecewise linear approximations to the Brownian motion) show that (4.30) is correct. A short discussion on the realization problem is included in section V.

References

[1] W.M. Wonham, “Some applications of stochastic differential equations to optimal non-linear filtering”, Technical Report 64-3, Feb. 1964, RIAS, Baltimore, Maryland.

[2] J.L. Doob Stochastic Processes, 1953.

[3] K. Ito, “On a formula concerning stochastic differentials”, Nagoya Math. J., Vol 3, 1951, pp. 55–65.

[4] K. Ito, “Lectures on Stochastic Processes”, Tata Inst. of Fundamental Research, Bombay, 1961, Chapter 5, pp. 172–210.

[5] E. Wong and M. Zakai, “On the relation between ordinary and stochastic differential equations”, ERL Report 64-26 (U. of California, Berkeley) to be published in Int’l. J. of Eng. Science (1965).

[6] E. Wong and M. Zakai, “The oscillation of stochastic integrals”, ARL Res. Rpt. No. 424, to be published in Z. fur Wahrscheinlichkeitstheorie (1965).

[7] E. Wong and M. Zakai, “On the convergence of ordinary integrals to stochastic integrals”, ARL Res. Rpt. No. 439, to be published in Annals Math. Stat. (1965)