INFORMATION AND CONTROL **11**, 102–137 (1967)

Infinite-Dimensional Filtering: The Kalman–Bucy Filter in Hilbert Space*

PETER L. FALB†

Information and Control Engineering, The University of Michigan, Ann Arbor, Michigan

1. INTRODUCTION

We examine the question of determining the "best" linear filter, in an expected squared error sense, for a signal generated by stochastic linear differential equation on a Hilbert space. Our results, which extend the development in Kalman and Bucy (1960), *rely heavily on the integration theory for Banach-space-valued functions of* Dunford and Schwartz (1958). In order to derive the Kalman–Bucy filter, we also need to define and discuss such concepts as stochastic process, covariance, orthogonal increments, Wiener process, and stochastic integral in a Hilbert space context. We do this making extensive use of the ideas in Doob (1953).

The two crucial points in our treatment are (1) our definition of the covariance as a bounded linear transformation, and (2) our use of a Fubini-type theorem involving the interchange of stochastic and Lebesgue integration. As a byproduct, we also obtain a fully rigorous theory for the finite-dimensional case which does not rely on Ito's Lemma (cf. Kushner, 1964). This is of some independent interest.

The remainder of the paper is divided into the following sections: 2. Preliminaries; 3. Wiener Processes; 4. Stochastic Integration; 5. An Existence Theorem; 6. The Wiener-Hopf Equation; 7. The Optimal Filter; 8. Concluding Remarks.

We introduce some basic preliminary notions, the most important of which is the covariance of two Hilbert-space-valued random variables, in Section 2. Then, we discuss Wiener processes and construct an infinite-

* This research was sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant Nr. AFOSR 814-66.

† Present Address: Brown University, Providence, Rhode Island.

dimensional example of such a process. In Section 4, we define stochastic integrals and develop some of their properties. Next, we prove a basic existence theorem for linear stochastic differential equations using some ideas of Beutler (1963). Having developed the necessary machinery, we state the filtering problem and prove a theorem (that involves a Wiener–Hopf equation) giving the basic necessary and sufficient condition for a solution of this problem in Section 6. We derive the various equations describing the optimal filter in Section 7 and make some final brief comments in Section 8.

2. PRELIMINARIES

We let $(\Omega, \mathcal{P}, \mu)$ be a probability space with $\mathcal{P}$ as Borel field and μ as measure throughout the paper. We recall (see Dunford and Schwartz, 1958) that, if X is a Banach space, then a function $x(\cdot)$ defined on Ω with values in X is measureable if (1) $x(\cdot)$ is essentially separably valued, and (2) $x^{-1}(\mathcal{O}) \in \mathcal{P}$ for each open set $\mathcal{O}$ in X. We shall call such measureable functions, *random variables*. If $x(\cdot)$ is an integrable random variable, then we say that $x(\cdot)$ has an expected value (or mean) which we denote by $E\{x(\cdot)\}$; i.e.,

$$E\{x(\cdot)\} = \int_\Omega x(\omega)\, d\mu. \tag{1}$$

We observe that $E\{x(\cdot)\}$ is also an element of the Banach space X. We shall assume from now on that the random variables which we consider have expected values and we shall often delete the explicit ω-dependence of these random variables.

In the sequel, we shall consider parameterized families of random variables, i.e., stochastic processes. In other words, we have:

DEFINITION 2.1. Let T be a real interval and let X be a Banach space. Then a function $x(t, \omega)$ from $T \times \Omega$ into X, which is measureable in the pair (t, ω) using Lebesgue measure on T, is called a stochastic process.

This definition is somewhat more restrictive than the usual one (cf. Doob, 1953) but is adequate for our purposes. Also, we often write $x(t)$ in place of $x(t, \omega)$ when discussing stochastic processes.

If X is a Banach space with dual X^*, then we write $\langle y^*, x\rangle$ for the operation of an element y^* of X^* on an element x of X. Now if $x_1 \in X$ and $y_1 \in X^*$, then we can define a mapping $x_1 \circ y_1^*$ of X into itself by setting

$$(x_1 \circ y_1^*)\, x = x_1 \langle y_1^*, x\rangle \tag{2}$$

for all $x \in X$. We observe that, if $X = R_n$ (i.e., X is finite-dimensional), then $y_1 \circ y_1^*$ can be identified with the matrix $x_1 y_1'$. Moreover, we have:

PROPOSITION 2.2 *Let X be a Banach space and let ψ be the mapping of $X \oplus X^*$ into $\mathfrak{L}(X, X)$ defined by*

$$\psi(x_1, y_1^*) = x_1 \circ y_1^*. \tag{3}$$

Then ψ has the following properties:

(a) *$\psi(x_1, y_1^*)$ is a linear transformation of X into X for all x_1 and y_1^*;*

(b) *ψ is continuous since*

$$\| \psi(x_1, y_1^*) \| = \| x_1 \| \, \| y_1^* \| \tag{4}$$

[Observe that this also implies that the linear transformation $\psi(x_1, y_1^*)$ is actually an element of $\mathfrak{L}(X, X)$.];

(c) *ψ is linear in both x_1 and y_1^*;*

(d) *the adjoint $\psi(x_1, y_1^*)^*$ of $\psi(x_1, y_1^*)$ is $y_1^* \circ x_1^{**}$, where x_1^{**} is x_1 viewed as an element of X^{**}, and hence, if X is reflexive, then $\psi(x_1, y_1^*)^* = y_1^* \circ x_1$.*

Proof. Properties (a) and (c) are obvious. As for (b), we simply note that

$$\begin{aligned} \| \psi(x_1, y_1^*) \| &= \sup_{\|x\| \le 1} \| (x_1 \circ y_1^*)\, x \| = \sup_{\|x\| \le 1} \| x_1 \langle y_1^*, x\rangle \| \\ &= \| x_1 \| \sup_{\|x\| \le 1} | \langle x_1^*, x\rangle | = \| x_1 \| \, \| y_1^* \| \end{aligned} \tag{5}$$

(see, for example, Kantorovich and Akilov, 1964). To establish (d), we note that if y^* is an element of X^*, then $\psi(x_1, y_1^*)^* (y^*)$ is an element of X^* and

$$\begin{aligned} \langle \psi(x_1, y_1^*)^*(y^*), x\rangle &= \langle y^*, \psi(x_1, y_1^*)x\rangle = \langle y^*, x_1 \langle y_1^*, x\rangle\rangle \\ &= \langle y^*, x_1\rangle \langle y_1^*, x\rangle = \langle y_1^* \langle y^*, x_1\rangle, x\rangle \\ &= \langle y_1^* \langle x_1^{**}, y^*\rangle, x\rangle = \langle (y_1^* \circ x_1^{**})(y^*), x\rangle, \end{aligned} \tag{6}$$

from which (d) follows.

COROLLARY 2.3 *If X is a Hilbert space (so that X^* can be identified with X), then $\psi(x_1, x_1)$ is symmetric and, if x is an element of X, then*

$$\langle x, \psi(x_1, x_2)\, x\rangle = \langle x, x_1\rangle \langle x, x_2\rangle \tag{7}$$ [1]

[1] Here $\langle\ ,\ \rangle$ denotes the inner product on X.

which may also be written as

$$\langle x^*, (x_1 \circ x_2)\, x\rangle = \langle x^*, x_1\rangle \langle x^*, x_2\rangle \tag{8}$$

(*where* x^* *is* x *viewed as an element of* X^*).

Proposition 2.2 leads us to the following:

PROPOSITION 2.4 *Let* $x(\bullet)$, $y^*(\bullet)$ *be* X *and* X^* *valued random variables, respectively. Then* $x(\bullet) \circ y^*(\bullet)$ *is an* $\mathcal{L}(X, X)$*-valued measureable function.*

Proof. We want to show that $x(\bullet) \circ y^*(\bullet)$ is measureable. Since $x(\bullet)$ and $y^*(\bullet)$ are essentially separably valued, it is clear that $x(\bullet) \circ y^*(\bullet)$ is essentially separably valued and so, it will be enough to prove that, if $\mathcal{O}$ is open in $\mathcal{L}(X, X)$, then $\{x(\bullet) \circ y^*(\bullet)\}^{-1}(\mathcal{O})$ is in $\mathcal{P}$ (see Dunford and Schwartz, 1958). Now,

$$\{\omega:\psi(x(\omega),\, y^*(\omega)) \in \mathcal{O}\} = \{\omega:(x(\omega),\, y^*(\omega)) \in \psi^{-1}(\mathcal{O})\} \tag{9}$$

But ψ is continuous implies that $\psi^{-1}(\mathcal{O})$ is open in $X \oplus X^*$; thus $\psi^{-1}(\mathcal{O}) = \bigcup \mathcal{O}_i \times \mathcal{O}_i^*$ where the $\mathcal{O}_i$, $\mathcal{O}_i^*$ are open in X and X^*, respectively. It follows that

$$\begin{aligned}\{\omega:(x(\omega), y^*(\omega)) \in \psi^{-1}(\mathcal{O})\}& \\ &= \bigcup\, [\{\omega:x(\omega) \in \mathcal{O}_i\} \cap \{\omega:y^*(\omega) \in \mathcal{O}_i^*\}] \\ &= \bigcup\, x^{-1}(\mathcal{O}_i) \cap y^{*-1}(\mathcal{O}_i^*)\end{aligned} \tag{10}$$

Since $x(\bullet)$ and $y^*(\bullet)$ are measureable, the proposition follows.

We now have:

DEFINITION 2.5. Let x, y^* be X and X^* valued random variables, respectively.[2] Then the covariance of x and y^*, in symbols cov $[x, y^*]$ is the element (if it exists) of $\mathcal{L}(X, X)$ given by

$$\operatorname{cov}\,[x, y^*] = E\{x \circ y^*\} - E\{x\} \circ E\{y^*\}. \tag{11}$$

If cov $[x, y^*] = \mathbf{0}$, then we say that x and y^* are uncorrelated. If X is a Hilbert space (so that X^* is identified with X) and if x and y are X valued random variables, then we write cov $[x, y]$ and speak of "the covariance of x and y."

The notion of covariance will play a crucial role in the sequel.

Now let us suppose that $X = H$ is a Hilbert space. Then $L^2(\Omega, H)$ is

[2] Note that we are deleting the ω dependence in accordance with our earlier remarks.

also a Hilbert space with inner product given by

$$\langle x, y \rangle_2 = \int_\Omega \langle x(\omega), y(\omega) \rangle \, d\mu, \tag{12}$$

where $\langle \;\; , \;\; \rangle$ denotes the inner product on H. We observe that if x and y are elements of $L^2(\Omega, H)$, then $E\{x \circ y\}$ exists so that $x \circ y$ is a random variable and cov $[x, y]$ exists. Moreover, we note that it is thus possible to speak of "wide sense" concepts in a straightforward manner. For example, we have:

DEFINITION 2.6. Let $x(t)$ be an H-valued random process. If $E\{\| x(t) \|^2\} < \infty$ for all t (i.e., $x(t) \in L^2(\Omega, H)$ for all t), then $x(t)$ is called a wide-sense martingale if[3]

$$\hat{E}\{x(t) \mid x(r), \; r \leqq s\} = x(s) \tag{13}$$

whenever $s < t$ where $\hat{E}\{x(t) \mid x(r), \; r \leqq s\}$ denotes the projection of $x(t)$ on the subspace of $L^2(\Omega, H)$ generated by the $x(r)$, $r \leqq s$.

We remark that it is also possible to introduce the notion of a martingale for H-valued random processes; however, this notion is not needed here and involves some complicated measure theoretic considerations which would take us too far afield from our main purpose.

We also require the following definition.

DEFINITION 2.7. Let $x(t)$ be an H-valued random process. Then $x(t)$ is called a process with orthogonal increments if

(1) $E\{\| [x(t) - x(s)] \circ [x(t) - x(s)] \|\} < \infty$ *for* $s, t \in T$;[4]

(2) $E\{[x(t_2) - x(s_2)] \circ [x(t_1) - x(s_1)]\} = \mathbf{0}$ *for* $s_1 < t_1 \leqq s_2 < t_2$.

If $E\{[x(t) - x(s)] \circ [x(t) - x(s)]\}$ depends only on $t - s$, then $x(t)$ is said to have (wide sense) stationary increments (see Doob, 1953).

To indicate some of the structure of processes with orthogonal increments, we now state and prove a simple proposition.

PROPOSITION 2.8 *Let $x(t)$ be a process with orthogonal increments. Then:*

(a) $E\{[x(t_1) - x(s_1)] \circ [x(t_2) - x(s_2)]\} = \mathbf{0}$ *for* $s_1 < t_1 \leqq s_2 < t_2$;

(b) *if* $\lambda(t)$ *is given by*

[3] This is to be interpreted in an almost everywhere sense.

[4] Note that, in view of (4), this is equivalent to $[x(t) - x(s)]$ being an element of $L^2(\Omega, H)$, i.e., $E\{\| x(t) - x(s) \|^2\} < \infty$.

$$\lambda(t) = \begin{cases} E\{[x(t) - x(t_0)] \circ [x(t) - x(t_0)]\}, & t \geqq t_0, \\ -E\{[x(t) - x(t_0)] \circ [x(t) - x(t_0)]\}, & t < t_0, \end{cases} \tag{14}$$

then

$$E\{[x(t) - x(s)] \circ [x(t) - x(s)]\} = \lambda(t) - \lambda(s), s < t; \tag{15}$$

(c) *if* $s < t$, *then* $\lambda(t) - \lambda(s)$ *is a positive element of* $\mathcal{L}(H, H)$ [*in this sense*, $\lambda(\cdot)$ *is a monotone-nondecreasing function*].

Proof. To establish (a), we simply note that, in view of (d) of Proposition 2.2,

$$\begin{aligned} &E\{[x(t_1) - x(s_1)] \circ [x(t_2) - x(s_2)]\} \\ &\qquad = E\{([x(t_2) - x(s_2)] \circ [x(t_1) - x(s_1)])^*\} \qquad (16)[5] \\ &\qquad = (E\{[x(t_2) - x(s_2)] \circ [x(t_1) - x(s_1)]\})^*. \end{aligned}$$

A simple calculation, which is omitted, verifies (b). As for (c), we note, first of all, that $\lambda(t) - \lambda(s)$ is symmetric by virtue of (d) of Proposition 2.2. Now, if $h \in H$, then

$$\begin{aligned} \langle[\lambda(t) - \lambda(s)]h, h\rangle &= \langle E\{[x(t) - x(s)] \circ [x(t) - x(s)]\}h, h\rangle \\ &= \langle E\{[x(t) - x(s)]\, \langle[x(t) - x(s)], h\rangle\}, h\rangle \qquad (17) \\ &= E\{\langle[x(t) - x(s)], h\rangle \cdot \langle[x(t) - x(s)], h\rangle\} \\ &\geqq 0, \end{aligned}$$

so that $\lambda(t) - \lambda(s)$ is positive.

Now if $x(t)$ is a process with orthogonal increments, then $\lambda(t)$ can be used to define an $\mathcal{L}(H, H)$-valued measure and it would be possible to develop a stochastic integral based on this measure (cf. Doob, 1953). However, for our purposes here, we shall be considerably less general and shall deal with processes which are analogous to Wiener processes. We define and study these processes in the next section.

3. WIENER PROCESSES

In this section, we introduce the notion of a Wiener process and exhibit an infinite-dimensional example of such a process. Our development of the stochastic integral will be based on Wiener processes. Now we have:

[5] Since $A(\cdot) \in L(\Omega, \mathcal{L}(H, H))$ implies that $(\int A(\omega)d\mu)^* = \int A^*(\omega)d\mu$ (cf. Dunford and Schwartz, 1958).

DEFINITION 3.1 Let $w(t)$ be an H-valued random process. Then $w(t)$ is called a Wiener process if the following conditions are satisfied:

(1) $E\{w(t)\} = 0$ for all t;

(2) $w(t)$ has orthogonal increments;

(3) $w(t)$ is continuous almost everywhere with respect to ω;

(4) $$E\{[w(t) - w(s)] \circ [w(t) - w(s)]\} = |t - s| W, \tag{18}$$

where W is a positive definite element of $\mathcal{L}(H, H)$ with $W\epsilon_\alpha = \lambda_\alpha\epsilon_\alpha$ for some orthonormal basis $\{\epsilon_\alpha\}$ of H; and,

(5) $$E\{\langle w(t_2) - w(s_2), S[w(t_1) - w(s_1)]\rangle\} = 0 \tag{19}$$

whenever $s_1 < t_1 \leqq s_2 < t_2$ and $S \in \mathcal{L}(H, H)$.[6]

The first question that comes to mind is: does there exist a Wiener process which is not finite-dimensional? We now show that the answer to this question is YES!

Let l_2 denote the Hilbert space of square-summable sequences and let $\epsilon_1, \epsilon_2, \cdots$ denote an orthonormal basis of l_2. If $b_i(t)$, $i = 1, 2, \cdots$ are independent Brownian motion processes with zero means and unit standard deviations, then we can consider the function

$$w(t) = \sum_{n=1}^{\infty} b_n(t)\epsilon_n/n^3 \tag{20}$$

viewed, for the moment, as a formal sum. We shall show that $w(t)$ can be reviewed as an l_2-valued Wiener process. To begin with, we have:

LEMMA 3.2 *Let* $w_N(t)$ *be given by*

$$w_N(t) = \sum_{n=1}^{N} b_n(t)\epsilon_n/n^3 \tag{21}$$

for $N = 1, 2, \cdots$. *Then*

(1) $w_N(t)$ *is an* l_2*-valued random variable with zero mean for all t and* $N = 1, 2, \cdots$;

(2) $w_N(t)$ *is continuous in t almost everywhere with respect to* ω *for* $N = 1, 2, \cdots$;

(3) $w_N(t)$ *is an element of* $L^2(\Omega, l_2)$ *for* $N = 1, 2, \cdots$;

(4) *the sequence* $w_N(t)$ *converges almost everywhere with respect to* ω *and this convergence is uniform in t for t in any interval* $[0, t_1]$;

[6] Equation (19) is an "independence" condition which is redundant in the finite dimensional case.

(5) *the sequence* $w_N(t)$ *converges in* $L^2(\Omega, l_2)$ *and this convergence is uniform in* t *for* t *in any interval* $[0, t_1]$.

Proof. The properties (1), (2), and (3) are immediate consequences of the definition of $w_N(t)$ and the properties of the Brownian motions $b_i(t)$. Now let

$$P_n = \{\omega : | b_n(t, \omega) |^2 > n^2\} \tag{22}$$

for t in $[0, t_1]$ and let $p_n = \mu(P_n)$. Then, as is well known (see Doob, 1953),

$$p_n \leqq c(t_1)/n^2 \tag{23}$$

and hence,

$$\sum_{n=1}^{\infty} p_n \leqq \sum_{n=1}^{\infty} c(t_1)/n^2 < \infty. \tag{24}$$

It follows from the Borel–Cantelli lemma that, if $\Lambda = \{\omega : \omega \epsilon$ infinitely many $P_n\}$, then $\mu(\Lambda) = 0$. If E_N is given by

$$E_N = \bigcup_{n=N}^{\infty} \{\omega : | b_n(t, \omega) |^2 \leqq n^2\} = \bigcup_{n=N}^{\infty} [\Omega - P_n] \tag{25}$$

for t in $[0, t_1]$, then

$$\Omega - \Lambda = \bigcup_{N=1}^{\infty} E_N \subseteq \Omega, \tag{26}$$

and so, $\mu(\bigcup_{N=1}^{\infty} E_N) = 1$. However, since $E_N \subseteq E_M$ when $M \geqq N$, and since

$$\| w_{N_2}(t, \omega) - w_{N_1}(t, \omega) \|^2 \leqq \sum_{N_1}^{N_2} \frac{| b_n(t, \omega) |^2}{n^6} \leqq c(t_1) \sum_{N_1}^{N_2} \frac{1}{n^2}, \tag{27}$$

where $\omega \epsilon E_N$, $N_2 \geqq N_1 \geqq N$, and t is in $[0, t_1]$, we can see that $\omega \epsilon \bigcup_1^{\infty} E_N$ implies that $w_N(t, \omega)$ converges and that this convergence is uniform in t. Combining this observation with (26), we immediately deduce (4). As for (5), we simply observe that

$$\int_{\Omega} \| w_N(t) - w_M(t) \|^2 \, d\mu \leqq \sum_{M}^{N} \int_{\Omega} \frac{| b_n(t) |^2}{n^6} \, d\mu \leqq \sum_{M}^{N} \frac{1}{n^6} \tag{28}$$

since $E\{b_n(t)^2\} = 1$. This completes the proof of the lemma.

Now, suppose that $\hat{w}(t)$ denotes the limit of the sequence $w_N(t)$ in

$L^2(\Omega, l_2)$ and that $w(t)$ denotes the limit of the sequence $w_N(t)$ in the almost-everywhere sense. Then, we claim that $\hat{w}(t) = w(t)$ almost everywhere with respect to Ω. To verify this claim, we note that the convergence of $w_N(t)$ to $\hat{w}(t)$ in $L^2(\Omega, l_2)$ implies that there is a subsequence $w_{N_j}(t)$ which converges to $\hat{w}(t)$ almost everywhere (see Dunford and Schwartz, 1958)[7]; but, $w_{N_j}(t)$ converges to $w(t)$ almost everywhere and so, $w(t) = \hat{w}(t)$ almost everywhere. Thus, $w(t)$, which can be written in the form of the sum (20), is an l_2-valued random variable with zero mean. Moreover, by (4), $w(t)$ is continuous almost everywhere with respect to ω. Also, since $w(t)$ and $w(s)$ are elements of $L^2(\Omega, l_2)$, we have $E\{\| w(t) - w(s) \|^2\} < \infty$ and so, in order to show that $w(t)$ has orthogonal increments, we must show that

$$E\{[w(t_2) - w(s_2)] \circ [w(t_1) - w(s_1)]\} = \mathbf{0} \tag{29}$$

for $s_1 < t_1 \leqq s_2 < t_2$. To do this, it will be sufficient to show that

$$E\{[w(t_2) - w(s_2)] \circ [w(t_1) - w(s_1)]\} \epsilon_n = 0 \tag{30}$$

for all the elements ϵ_n of our orthonormal basis of l_2. But

$$\begin{aligned}
&E\{[w(t_2) - w(s_2)] \circ [w(t_1) - w(s_1)]\} \epsilon_n \\
&\quad = E\{[w(t_2) - w(s_2)]\langle w(t_1) - w(s_1), \epsilon_n\rangle\} \\
&\quad = E\{[w(t_2) - w(s_2)]\,(b_n(t_1) - b_n(s_1))/n^3\} \\
&\quad = \sum_{m=1}^{\infty} E\{(b_m(t_2) - b_m(s_2))(b_n(t_1) - b_n(s_1)) \cdot\} \epsilon_{m/m^3n^3},
\end{aligned} \tag{31}$$

by virtue of Eq. (20). Since the $b_i(t)$ are independent Brownian-motion processes, we have established (30). Thus, we have verified that $w(t)$ satisfies conditions (1)–(3) of Definition 3.1.

Now, we observe that, by the definition of $w(t)$ [see Eq. (20)] and the fact that the $b_i(t)$ are independent Brownian-motion processes,

$$\begin{aligned}
&E\{[w(t) - w(s)] \circ [w(t) - w(s)]\} \epsilon_\alpha \\
&\qquad = E\{(b_\alpha(t) - b_\alpha(s))^2\} \frac{\epsilon_\alpha}{\alpha^6} = | t - s | \frac{\epsilon_\alpha}{\alpha^6}.
\end{aligned} \tag{32}$$

Hence, if we define W by setting $W\epsilon_\alpha = \dfrac{\epsilon_\alpha}{\alpha^6}$ and extending by linearity, then we deduce from (32) that

[7] The subsequence may depend on t.

$$E\{[w(t) - w(s)] \circ [w(t) - w(s)]\} = |t - s| W. \tag{33}$$

Since W is clearly positive-definite, we have verified that condition (4) of Definition 3.1 is satisfied by $w(t)$. So, all that remains is to show that condition (5) of Definition 3.1 holds for $w(t)$. This we do in the following lemma:

LEMMA 3.3 *Let S be an element of $\mathfrak{L}(H, H)$. Then*

$$E\{\langle w(t_2) - w(s_2), S[w(t_1) - w(s_1)]\rangle\} = 0 \tag{34}$$

whenever $s_1 < t_1 \leqq s_2 < t_2$.

Proof. To simplify notation, let us set

$$\begin{aligned} \Delta_1 &= w(t_1) - w(s_1), & \Delta_2 &= w(t_2) - w(s_2), \\ \Delta_{1N} &= w_N(t_1) - w_N(s_1), & \Delta_{2N} &= w_N(t_2) - w_N(s_2), \end{aligned} \tag{35}$$

where $w_N(t)$ is given by (21). Then

$$\langle \Delta_1, S\Delta_2 \rangle = \langle \Delta_1 - \Delta_{1N}, S\Delta_2 \rangle + \langle \Delta_{1N}, S[\Delta_2 - \Delta_{2N}] \rangle + \langle \Delta_{1N}, S\Delta_{2N} \rangle \tag{36}$$

and, hence,

$$\begin{aligned} E\{\langle \Delta_1, S\Delta_2 \rangle\} = {} & E\{\langle \Delta_1 - \Delta_{1N}, S\Delta_2 \rangle\} \\ & + E\{\langle \Delta_{1N}, S[\Delta_2 - \Delta_{2N}] \rangle\} + E\{\langle \Delta_{1N}, S\Delta_{2N} \rangle\}. \end{aligned} \tag{37}$$

Now[8]

$$\begin{aligned} | E\{\langle \Delta_1 - \Delta_{1N}, S\Delta_2 \rangle\} | & \leqq E\{| \langle \Delta_1 - \Delta_{1N}, S\Delta_2 \rangle |\} \\ & \leqq E\{\| \Delta_1 - \Delta_{1N} \| \cdot \| S \| \cdot \| \Delta_2 \|\} \\ & \leqq \| \Delta_1 - \Delta_{1N} \|_2 \| S \| \cdot \| \Delta_2 \|_2 \end{aligned} \tag{38}$$

and hence, in view of Lemma 3.2, we have $E\{\langle \Delta_1 - \Delta_{1N}, S\Delta_2 \rangle\} = 0$. In a similar way, we can show that $E\{\langle \Delta_{1N}, S[\Delta_2 - \Delta_{2N}] \rangle\} = 0$. As for the third term in (37), we have

$$\begin{aligned} \langle \Delta_{1N}, S\Delta_{2N} \rangle = \sum_{m=1}^{N} \sum_{n=1}^{N} & (b_m(t_2) - b_m(s_2))(b_n(t_1) \\ & - b_n(s_1)) \; \langle \epsilon_m, S\epsilon_n \rangle / m^3 n^3. \end{aligned} \tag{39}$$

Since the $b_i(t)$ are independent Brownian-motion processes, it follows that $E\{\langle \Delta_{1N}, S\Delta_{2N} \rangle\} = 0$. This completes the proof of the lemma.

[8] $\| \;\|_2$ denotes the $L^2(\Omega, l_2)$ norm and the last inequality in (38) follows from Holder's inequality (see Dunford and Schwartz, 1958).

We have thus shown that $w(t)$ is an l_2-valued Wiener process and so have constructed an example of an infinite-dimensional Wiener process.

4. STOCHASTIC INTEGRATION

We are now prepared to discuss the stochastic integral and to develop some properties of this integral; this will enable us to discuss stochastic differential equations and to state the filtering problem. Our definition of the integral will be based on Definition 3.1. So let us suppose that H is a Hilbert space, that $w(t)$ is an H-valued Wiener process, and that $S(t)$ is a step function from T into $\mathfrak{L}(H, H)$. In other words, there are $t_0 < t_1 < \cdots < t_n$ in T and S_j, $j = 1, 2, \cdots, n$ in $\mathfrak{L}(H, H)$ such that

$$S(t) = \begin{cases} \mathbf{0} & t < t_0 \\ S_j & t_{j-1} \leqq t < t_j \qquad j = 1, 2, \cdots, n \\ \mathbf{0} & t \geqq t_n \end{cases}$$

Then, the stochastic integral of $S(t)$ with respect to $dw(t)$, in symbols: $\int S(t)\, dw(t)$, is defined by

$$\int S(t)\, dw(t) = \sum_{j=1}^{n} S_j[w(t_j) - w(t_{j-1})]. \tag{41}$$

We observe that this integral is an H-valued random variable with zero mean.[9] We shall "build up" the stochastic integral by an approximation procedure based on (41). Toward that end, we have

LEMMA 4.1 *Let $R(t)$ and $S(t)$ be $\mathfrak{L}(H, H)$-valued step functions. Then*

$$E\left\{\left[\int R(t)\, dw(t)\right] \circ \left[\int S(t)\, dw(t)\right]\right\} = \int R(t) W S^*(t)\, dt \tag{42}$$

and

$$E\left\{\left\| \int R(t)\, dw(t) \right\| \cdot \left\| \int S(t)\, dw(t) \right\|\right\} \leqq \int \| R(t) \| \cdot \| W \| \cdot \| S(t) \|\, dt. \tag{43}$$

Proof. By adding $\mathbf{0}$'s if necessary, we may suppose that there are $t_0 < t_1 < \cdots < t_n$ in T such that

[9] Of course, the integral is to be viewed in an almost everywhere sense but we shall slur over this point here.

$$R(t) = \begin{cases} \mathbf{0}, & t < t_0 \\ R_j, & t_{j-1} \leqq t < t_j \\ \mathbf{0}, & t \geqq t_n, \end{cases} \qquad S(t) = \begin{cases} \mathbf{0} & t < t_0 \\ S_k & t_{k-1} \leqq t < t_k \\ \mathbf{0} & t \geqq t_n. \end{cases} \tag{44}$$

It then follows that

$$\left[\int R(t)\, dw(t)\right] \circ \left[\int S(t)\, dw(t)\right] = \sum_{j=1}^{n} \sum_{k=1}^{n} R_j[w(t_j) - w(t_{j-1})] \circ S_k[w(t_k) - w(t_{k-1})] \tag{45}$$

and hence that

$$E\left\{\left[\int R(t)\, dw(t) \circ\right]\left[\int S(t)\, dw(t)\right]\right\} = \sum_{j=1}^{n} \sum_{k=1}^{n} R_j E\{[w(t_j) - w(t_{j-1})] \circ [w(t_k) - w(t_{k-1})]\} S_k^* \tag{46}$$

$$= \sum_{j=1}^{n} R_j(t_j - t_{j-1}) W S_j^* = \int R(t) W S^*(t)\, dt \tag{47}$$

by virtue of the properties of $w(t)$. Similarly, we deduce from (45), Proposition 2.2, and the properties of $w(t)$ that

$$E\left\{\left\|\int R(t)\, dw(t)\right\| \cdot \left\|\int S(t)\, dw(t)\right\|\right\}$$

$$\leqq \sum_{j=1}^{n} \sum_{k=1}^{n} \| R_j \| \, \| E\{[w(t_j) - w(t_{j-1})] \circ [w(t_k) - w(t_{k-1})]\} \| \cdot \| S_k^* \| \tag{48}$$

$$\leqq \sum_{j=1}^{n} \| R_j \| \cdot \| W \| \cdot \| S_j^* \| \, (t_j - t_{j-1}) = \int \| R(t) \| \, \| W \| \, \| S(t) \| \, dt$$

as $\| S_j^* \| = \| S_j \|$ (see Dunford and Schwartz, 1958).

In view of Lemma 4.1, we can define a stochastic integral in a manner analogous to that used in Doob (1953). In other words, if $\Phi(t)$ is an element of $L^2(T, \mathcal{L}(H, H))$ so that

$$\int \| \Phi(t) \|^2 \, dt < \infty, \tag{49}$$

then $\Phi(t)$ is a limit of step functions $S_m(t)$; i.e.,

$$\lim_{m\to\infty} \int \|\Phi(t) - S_m(t)\|^2 \, dt = 0 \tag{50}$$

and we can define the *stochastic integral of* $\Phi(t)$ *with respect to* $dw(t)$, in symbols: $\int \Phi(t)\, dw(t)$, by letting

$$\int \Phi(t)\, dw(t) = \underset{m\to\infty}{\text{l.i.m.}} \int S_m(t)\, dw(t); \tag{51}$$

i.e.

$$\lim_{m\to\infty} E\left\{\left\| \int \Phi(t)\, dw(t) - \int S_m(t)\, dw(t) \right\|^2\right\} = 0 \tag{52}$$

so that $\int \Phi(t)\, dw(t)$ is the limit in mean square of the $\int S_m(t)\, dw(t)$. We note that (52) follows from (50) by virtue of the inequality (43). Moreover, $\int \Phi(t)\, dw(t)$ is also an H-valued random variable with zero mean (see Dunford and Schwartz, 1958). Now we have:

LEMMA 4.2 *Let* $\Phi(t)$ *and* $\Psi(t)$ *be elements of* $L^2(T, \mathfrak{L}(H, H))$. *Then*

$$\begin{aligned} &E\left\{\left\| \int \Phi(t)\, dw(t) \right\| \cdot \left\| \int \Psi(t)\, dw(t) \right\|\right\} \\ &\qquad\qquad \leqq \int \|\Phi(t)\| \cdot \|W\| \cdot \|\Psi(t)\|\, dt. \end{aligned} \tag{53}$$

Proof. Let $R_m(t)$ and $S_n(t)$ be sequences of step functions that define $\Phi(t)$ and $\Psi(t)$, respectively. Then, using Proposition 2.2,

$$\begin{aligned} &E\left\{\left\| \int \Phi(t)\, dw(t) \right\| \left\| \int \Psi(t)\, dw(t) \right\|\right\} \\ &\qquad\qquad = E\left\{\left\| \left[\int \Phi(t)\, dw(t)\right] \circ \left[\int \Psi(t)\, dw(t)\right] \right\|\right\}. \end{aligned} \tag{54}$$

But, again using Proposition 2.2,

$$\begin{aligned} &E\left\{\left\| \left[\int \Phi(t)\, dw(t)\right] \circ \left[\int \Psi(t)\, dw(t)\right] \right\|\right\} \\ &\qquad \leqq E\left\{\left\| \int [\Phi(t) - R_m(t)]\, dw(t) \right\| \cdot \left\| \int \Psi(t)\, dw(t) \right\|\right\} \\ &\qquad + E\left\{\left\| \int R_m(t)\, dw(t) \right\| \cdot \left\| \int [\Psi(t) - S_m(t)]\, dw(t) \right\|\right\} \\ &\qquad + E\left\{\left\| \int R_m(t)\, dw(t) \right\| \cdot \left\| \int S_m(t)\, dw(t) \right\|\right\}. \end{aligned} \tag{55}$$

By virtue of Hölder's inequality, we have

$$\lim_{m\to\infty} E\left\{\left\|\int [\Phi(t) - R_m(t)]\, dw(t)\right\|\cdot\left\|\int \Psi(t)\, dw(t)\right\|\right\} = 0,$$
$$\lim_{m\to\infty} E\left\{\left\|\int R_m(t)\, dw(t)\right\|\cdot\left\|\int [\Psi(t) - S_m(t)]\, dw(t)\right\|\right\} = 0, \tag{56}$$

and so, in view of (43), we need only show that

$$\lim_{m\to\infty} \int \| R_m(t) \|\cdot\| W \|\, \| S_m(t) \|\, dt = \int \| \Phi(t) \|\cdot\| W \|\, \| \Psi(t) \|\, dt, \tag{57}$$

or, equivalently, that

$$\lim_{m\to\infty} \int \| R_m(t) \|\cdot\| S_m(t) \|\, dt = \int \| \Phi(t) \|\cdot\| \Psi(t) \|\, dt. \tag{58}$$

But

$$\left| \int \| \Phi(t) \|\, \| \Psi(t) \|\, dt - \int \| R_m(t) \|\cdot\| S_m(t) \|\, dt \right|$$
$$\leqq \int \| \Phi(t) - R_m(t)\|\cdot\| \Psi(t) \|\, dt \qquad (59)^{10}$$
$$+ \int \| R_m(t) \|\cdot\| \Psi(t) - S_m(t) \|\, dt.$$

Combining (59) with Holder's inequality, we obtain (58) and thus establish the lemma.

The argument used in proving Lemma 4.2 is quite standard and will be used in various guises in the sequel. In the interests of economy of exposition, we shall not repeat the argument but shall simply use the phrase "by an approximation argument."

Now we observe that if $T = [t_0, t_1]$ (or $[t_0, \infty)$), then

$$x(t) = \int_{t_0}^{t} \Phi(s)\, dw(s), \qquad t \in T \tag{60}$$

may be viewed as an H-valued random process. If $\Phi(s)$ is a step function, then $\int_{t_0}^{t} \Phi(s)\, dw(s)$ has a version which is measureable in the pair (t, ω) in view of the continuity of the sample paths of $w(\cdot)$. We show that this is true for the general case in Theorem 4.10. Assuming this for the

[10] Since $| \| A \| - \| B \| | \leqq \| A - B \|$.

moment, we now show that $x(t)$ is a wide-sense martingale (see Definition 2.6).[11]

PROPOSITION 4.3. *Let $x(t)$ be the stochastic integral of* (60). *Then $x(t)$ is an element of $L^2(\Omega, H)$ and $x(t_2) - x(s_2)$ is orthogonal to $x(t)$ for $t \leqq s_2 < t_2$ in the $L^2(\Omega, H)$ sense; i.e.,*

$$E\{\langle x(t_2) \;- x(s_2),\;\; x(t)\rangle\} \;=\; 0 \tag{61}$$

Proof. The assertion that $x(t)$ is in $L^2(\Omega, H)$ is an immediate consequence of the definition of the stochastic integral. (61) is established by "an approximation argument" which is omitted.

COROLLARY 4.4. $E\{\| x(s) \|^2\} \leqq E\{ \| x(t) \|^2\}$ *if* $s \leqq t$. [For a proof of the corollary see Doob, 1953, p. 165.]

We shall in the sequel have occasion to consider double integrals of the form

$$\int\int \Phi(s, t)\, ds\, dw(t), \tag{62}$$

where the $\mathfrak{L}(H, H)$-valued function $\Phi(s, t)$ is measureable in s and continuous in t; we shall define this integral in terms of iterated integrals. (cf. Doob, 1953). We begin with a lemma.

LEMMA 4.5. *Suppose that $\Phi(s, t)$ is measureable in s and t and that*

$$\int_T \| \Phi(s, t) \|^2\, dt < \infty \tag{63}$$

almost everywhere in s. Then the stochastic integral

$$y(s) = \int_T \Phi(s, t)\, dw(t) \tag{64}$$

can be defined in such a way that the process $y(s)$ is measureable.

Proof. Suppose first of all that

$$\Phi(s,\; t) \;=\; \sum_j \;\Phi_{1j}(s)\Phi_{2j}(t), \tag{65}$$

where the $\Phi_{1j}(s)$ and $\Phi_{2j}(t)$ are measureable and

$$\int_T \| \Phi_{2j}(t) \|^2\, dt < \infty. \tag{66}$$

[11] As noted in Section 2, it is possible to introduce the notion of a martingale for H-valued random processes; with this notion in hand, we could show that $x(t)$ was a martingale. However, we shall not pursue this matter here.

Then, by the definition of the stochastic integral,

$$y(s) = \sum_j \Phi_{1j}(s) \int_T \Phi_{2j}(t)\, dw(t) \tag{67}$$

so that $y(s)$ is measureable in the pair (s, ω), being a finite sum of products of functions measureable in each variable. The lemma then follows by "an approximation argument."

We are now ready to define the double integral (62). Let E be an interval and suppose that $\Phi(s, t)$ is measureable in s and t. If

$$\int_T \left[\int_E \| \Phi(s, t) \| \, ds \right]^2 dt < \infty, \tag{68}$$

then

$$\Psi(t) = \int_E \Phi(s, t)\, ds \tag{69}$$

is an element of $L^2(T, \mathfrak{L}(H, H))$ and so $\int \Psi(t)\, dw(t)$ is defined; i.e., the iterated integral

$$\int_T \left[\int_E \Phi(s, t)\, ds \right] dw(t) = y_1(\omega) \tag{70}$$

is well-defined. Moreover,

$$E\{\| y_1 \|^2\} \leqq \int_T \| \Psi(t) \|^2 \cdot \| W \| \, dt < \infty. \tag{71}$$

On the other hand, if

$$\int_E \left[\int_T \| \Phi(s, t) \|^2 \, dt \right]^{1/2} ds < \infty, \tag{72}$$

then

$$\Theta(s) = \int_T \Phi(s, t)\, dw(t) \tag{73}$$

may be assumed measureable in the pair (s, ω) by virtue of Lemma 4.5 and $\int \Theta(s)\, ds$ is well-defined; i.e., the iterated integral

$$\int_E \left[\int_T \Phi(s, t)\, dw(t) \right] ds = y_2(\omega) \tag{74}$$

is well-defined. Moreover,

$$E\{\| y_2 \|\} \leq \int_E E\left\{\left\| \int_T \Phi(s, t)\, dw(t) \right\|\right\} ds$$
$$\leq \int_E \left[\int_T \| \Phi(s, t) \|^2\, dt\right]^{1/2} \| W \|^{1/2}\, ds < \infty. \tag{75}$$

We now have:

LEMMA 4.6 *Suppose that both* (68) *and* (72) *are valid. Then* $y_1(\omega) = y_2(\omega)$ *almost everywhere.*

Proof. Suppose first of all that

$$\Phi(s, t) = \sum_j \Phi_{1j}(s) \quad \Phi_{2j}(t) \tag{76}$$

with

$$\int_E \| \Phi_{1j}(s) \|\, ds < \infty, \qquad \left[\int_T \| \Phi_{2j}(t) \|^2\, dt\right]^{1/2} < \infty. \tag{77}$$

Then both (68) and (72) are valid and it is clear that $y_1(\omega) = y_2(\omega)$. The lemma then follows by "an approximation argument."

The substance of Lemma 4.6 is that the order of integration is immaterial provided that both (68) and (72) hold. We, of course, define the double integral (62) when either (68) or (72) are valid by means of a suitable iterated integral. We now have:

COROLLARY 4.7 (Fubini). *Suppose that* E *is finite and that* $\Phi(s, t)$ *is an element of* $L^2(E \times T, \mathfrak{L}(H, H))$ *i.e.,*

$$\iint \| \Phi(s, t) \|^2\, ds\, dt < \infty. \tag{78}$$

Then both (68) *and* (72) *are valid and hence, the order of integration in* (62) *is immaterial.*

Proof. Since E is finite, 1 is an element of $L^2(E, R)$. It then follows from Holder's inequality that

$$\int_T \left[\int_E \| \Phi(s, t) \|\, ds\right]^2 dt \leq \lambda(E)^2 \int_T \left[\int_E \| \Phi(s, t) \|^2\, ds\right] dt < \infty, \tag{79}$$

where $\lambda(E)$ is the Lebesgue measure of E. Thus (68) holds. On the other hand, if $f(s)$ is given by

$$f(s) = \int_T \| \Phi(s, t) \|^2\, dt, \tag{80}$$

then $f(s) \geqq 0$ and it will follow from (78) that $[f(s)]^{1/2}$ is an element of $L^2(E, R)$. In view of Holder's inequality, we then observe that $[f(s)]^{1/2} \cdot 1$ is an element of $L^1(E, R)$ and that

$$\int_E [f(s)]^{1/2}\, ds = \int_E \left[\int_T \| \Phi(s, t) \|^2\, dt\right]^{1/2} ds \\ \leqq [\lambda(E)]^{1/2} \int_E \int_T \| \Phi(s, t) \|^2\, dt\, ds < \infty \tag{81}$$

Thus (72) holds and the corollary is established.

COROLLARY 4.8 *Suppose that E and T are compact intervals. If $\Phi(s, t)$ is a regulated mapping of $E \times T$ into $\mathcal{L}(H, H)$*[12] *(in particular, if $\Phi(s, t)$ is continuous), then both* (68) *and* (72) *hold.*

We shall make frequent use of Corollaries 4.7 and 4.8 in what follows.

Now we observe that if $S(t)$ is a step function and if $T = [T_0, T_1]$ (or $[t_0, \infty)$), then the H-valued random process

$$y(t) = \int_{t_0}^{t} S(s)\, dw(s), \qquad t \in T \tag{82}$$

is continuous in t almost everywhere with respect to ω, and hence is Lebesgue-integrable[13] with respect to t on any finite subinterval of T almost everywhere with respect to ω. In the finite-dimensional case, it is reasonably easy to prove that the stochastic integral

$$x(t) = \int_{t_0}^{t} \Phi(s)\, dw(s) \tag{83}$$

is also continuous in t (see, for example, Doob, 1953) and therefore, Lebesgue-integrable. Here, in order to avoid considerable complexity relating to the notions of martingale and semimartingale for H-valued random processes, we shall not discuss the continuity of the "sample paths" of $x(t)$. However, we do have the following results:

LEMMA 4.9. *Let $S(t)$ be a step function and let $s < t$ be elements of T. Then*

$$E\left\{\left\| \int_{t_0}^{t} S(a)\, dw(a) - \int_{t_0}^{s} S(a)\, dw(a) \right\|^2\right\} \leqq c \cdot | t - s |, \tag{84}$$

[12] This means that $\Phi(s, t)$ is the limit of a uniformly convergent sequence of finite sums of products of step functions (cf. Dieudonne, 1960).

[13] In the sense of Dieudonne (1960).

where c is a constant (independent of s and t). The proof of this lemma is a simple calculation based on the properties of $w(\cdot)$ and is omitted.

THEOREM 4.10. *Let $\Phi(t)$ be an element of $L^2(T, \mathfrak{L}(H, H))$. Then $x(t, \omega) = \int_{t_0}^t \Phi(s)\, dw(s)$ is measureable in the pair (t, ω) and $\int x(t, \omega)\, dt$ exists almost everywhere with respect to ω on every finite subinterval of T.*

Proof. We shall apply Theorem 17, p. 198, of Dunford and Schwartz (1958). Let $F(t)$ be the mapping of T into $L^2(\Omega, H)$ given by

$$F(t) = x(t, \cdot) \tag{85}$$

We claim that $F(t)$ is continuous. Assuming for the moment that this claim is valid and that E is a finite subinterval of T, then, by the aforementioned Theorem 17, there is a function $f(t, \omega)$, measureable in the pair (t, ω), such that (1) $f(t, \cdot) = x(t, \cdot)$ almost everywhere in t; (2) $f(\cdot, \omega)$ is integrable in t almost everywhere in ω; and (3) $\int F(t)\, dt = \int f(t, \cdot)\, dt$. In view of the nonuniqueness of the stochastic integral,[14] the theorem will follow. Now let us verify the claim. Suppose that $s < t$ and that $S_N(a)$ is a step function. Then, letting

$$x_N(t) = \int_{t_0}^t S_N(a)\, dw(a), \qquad t \in T, \tag{86}$$

we have

$$\begin{aligned} \| F(t) - F(s) \|^2 &= E\{ \| x(t) - x(s) \|^2 \} \\ &\leqq E\{[\| x(t) - x_N(t) \| + \| x_N(t) - x_N(s) \| \\ &\qquad + \| x_N(s) - x(s) \|]^2\} \\ &\leqq 4E\{ \| x(t) - x_N(t) \|^2 \} + 4E\{ \| x_N(t) \\ &\qquad - x_N(s) \|^2\} + 4E\{ \| x_N(s) - x(s) \|^2\} \end{aligned} \tag{87}$$

[since $(| A | + | B |)^2 \leqq 2 | A |^2 + 2 | B |^2$]. Now let $\epsilon > 0$ be given; then, since $\Phi(t)$ is a limit of step functions in $L^2(T, \mathfrak{L}(H, H))$, we can choose an $S_N(a)$ such that

$$E\{ \| x(t) - x_N(t) \|^2\} < \epsilon/12, \qquad E\{ \| x_N(s) - x(s) \|^2\} < \epsilon/12 \tag{88}$$

as

$$\begin{aligned} E\{\| x(t) - x_N(t) \|^2\} &\leqq \int_{t_0}^t \| \Phi(a) - S_N(a) \|^2 \| W \|\, da \\ &\leqq \int_T \| \Phi(a) - S_N(a) \|^2 \cdot \| W \|\, da. \end{aligned} \tag{89}$$

[14] In other words, we may replace $x(t, \omega)$ by $f(t, \omega)$.

In view of lemma 4.9, it will follow that if (say)

$$|t - s| < \epsilon/12(c_N + 1) \tag{90}$$

where c_N is the constant of lemma 4.9 for $S_N(a)$, then

$$\| F(t) - F(s) \|^2 < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon \tag{91}$$

and hence, that $F(t)$ is continuous. This completes the proof of the theorem.

5. AN EXISTENCE THEOREM

We are now prepared to discuss stochastic linear differential equations on the Hilbert space H. So, if we suppose that x_0 is an H-valued random variable with $E\{ \| x_0\|^2\} < \infty$, that $A(t)$ is a regulated mapping of T into $\mathcal{L}(H, H)$, and that $M(t)$ is an element of $L^2(T, \mathcal{L}(H, H))$, then we can consider the following (stochastic) integral equation:

$$x(t) = x_0 + \int_{t_0}^{t} A(s)x(s)\, ds + \int_{t_0}^{t} M(s)\, dw(s). \tag{92}$$

We often write (92) in the form

$$dx = A(t)xdt + M(t)\, dw,\ x(t_0) = x_0 \tag{93}$$

and we speak of (93) as a "stochastic linear differential equation." Intuitively, we write (93) in the form

$$\dot{x} = A(t)x + M(t)\xi, \qquad x(t_0) = x_0 \tag{94}$$

where ξ is "white noise with covariance $W\delta(t - \tau)$." With regard to (92), we have:

THEOREM 5.1. *Let x_0 be an H-valued random variable with $E\{ \| x_0\|^2\} < \infty$. Let $A(t)$ be a regulated mapping from T into $\mathcal{L}(H, H)$ and let $M(t)$ be an element of $L^2(T, \mathcal{L}(H, H))$. Let $\Phi(t, t_0)$ be the fundamental linear transformation* (i.e., *resolvent*) *of the* (*nonstochastic*) *linear differential equation*

$$\dot{h} = A(t)h \tag{95}$$

on H (see Dieudonne, 1960). *Then* (92) *has a* (*essentially unique*) *solution $x(t)$ which is given by*

$$x(t) = \Phi(t, t_0)\left\{x_0 + \int_{t_0}^{t} \Phi(t_0, s)M(s)\, dw(s)\right\}. \tag{96}$$

Moreover, $E\{ \| x(t) \|^2\} < \infty$ and $E\{x(t)\} = \Phi(t, t_0)E\{x_0\}$.

Proof. Since (Dieudonne, 1960).

$$x_0 + \int_{t_0}^{t} A(s)\Phi(s, t_0)x_0\, ds = \Phi(t, t_0)x_0, \tag{97}$$

we may suppose, without loss of generality, that $x_0 = 0$. Thus, setting

$$y(s) = \int_{t_0}^{s} \Phi(t_0, \tau)M(\tau)\, dw(\tau) \tag{98}$$

and noting that $y(s)$ is an H-valued random process in view of Theorem 4.10 and the continuity of $\Phi(t_0, \tau)$,[15] we want to show that

$$\Phi(t, t_0)y(t) = \int_{t_0}^{t} A(s)\Phi(s, t_0)y(s)\, ds + \int_{t_0}^{t} M(s)\, dw(s). \tag{99}$$

We note that the Lebesgue-integral term in (99) may be viewed as an iterated integral in the following way: let $\Psi(s, \tau)$ be given by

$$\Psi(s, \tau) = \begin{cases} A(s)\Phi(s, t_0)\Phi(t_0, \tau)M(\tau), & \tau \leqq s, \\ 0 & \tau > s; \end{cases} \tag{100}$$

then

$$\int_{t_0}^{t} A(s)\Phi(s, t_0)y(s)\, ds = \int_{t_0}^{t} \left[\int_{t_0}^{t} \Psi(s, \tau)\, dw(\tau) \right] ds. \tag{101}$$

But, in view of Lemma 5.2 (which follows), $\Psi(s, \tau)$ is an element of $L^2(E \times E, \mathfrak{L}(H, H))$ where E is any finite subinterval of T of the form $[t_0, t_1]$ with $t_1 \geqq t$. We then deduce from Corollary 4.7 that

$$\begin{aligned} \int_{t_0}^{t} A(s)\Phi(s, t_0)y(s)\, ds \\ &= \int_{t_0}^{t} \left[\int_{t_0}^{t} \Psi(s, \tau)\, ds \right] dw(\tau) \\ &= \int_{t_0}^{t} \left[\int_{\tau}^{t} A(s)\Phi(s, t_0)\, ds \right] \Phi(t_0, \tau)M(\tau)\, dw(\tau). \end{aligned} \tag{102}$$

However,

$$A(s)\Phi(s, t_0) = d\Phi(s, t_0)/ds \tag{103}$$

and

$$\Phi(a, t_0)\Phi(t_0, b) = \Phi(a, b), \qquad \Phi(a, a) = I \tag{104}$$

[15] which implies that $\Phi(t_0, \tau)M(\tau)$ is in $L^2(E, \mathfrak{L}(H, H))$ for any finite subinterval E of T.

(see Dieudonne; 1960), so that

$$\begin{aligned}\int_{t_0}^{t} A(s)\Phi(s, t_0)y(s)\, ds &= \Phi(t, t_0) \int_{t_0}^{t} \Phi(t_0, \tau)M(\tau)\, dw(\tau) \\ &\quad - \int_{t_0}^{t} \Phi(\tau, \tau)M(\tau)\, dw(\tau) = \Phi(t, t_0)y(t) - \int_{t_0}^{t} M(\tau)\, dw(\tau).\end{aligned} \tag{105}$$

Thus the main assertion of the theorem is established. The final point that $E\{\| x(t) \|^2\} < \infty$ and $E\{x(t)\} = \Phi(t, t_0)\, E\{x_0\}$ is obvious.

LEMMA 5.2. *Let E be a compact interval and let X be a Banach space. Let $f(t)$ map E into $\mathcal{L}(X, X)$ and let $g(t)$ map E into $\mathcal{L}(X, X)$ (or X). If one of the maps f, g is in $L^p(E, \mathcal{L}(X, X))$ (or, in the case of g, $L^p(E, \mathcal{L}(X, X))$ or $L^p(E, X)$) and the other is regulated then $f(t)g(t)$ is an element of $L^p(E, \mathcal{L}(X, X))$ (or $L^p(E, X)$).*

Proof. For example, suppose that f is regulated and that g is in $L^p(E, \mathcal{L}(X, X))$. If f is a step function, then the result is clear. On the other hand, if f is regulated then (Dieudonne, 1960) there is a sequence f_N of step functions such that $f_N(\cdot)$ converges to $f(\cdot)$ uniformly on E. It follows that $f_N(t)g(t)$ converges to $f(t)g(t)$ almost everywhere and the $\| f_N(t)g(t) \| \leqq M \| g(t) \|$ for some $M > 0$ almost everywhere. But $Mg(t)$ is in $L^p(E, \mathcal{L}(X, X))$ and the lemma follows by the Lebesgue dominated convergence theorem (Dunford and Schwartz, 1958, p. 151). The other cases are treated in a similar manner.

COROLLARY 5.3. *Let $C(t)$ be a regulated mapping of E into $\mathcal{L}(X, X)$ and let $x(t)$ be generated by* (93). *Then*

$$y(t) = \int_{t_0}^{t} C(s)A(s)x(s)\, ds + \int_{t_0}^{t} C(s)M(s)\, dw(s) \tag{106}$$

is a well-defined H valued random process with $E\{\| y(t) \|^2\} < \infty$.

In view of Corollary 5.3, we shall often write $\int_{t_0}^{t} C(s)\, dx(s)$ in place of $y(t)$. In other words, we have

$$\int_{t_0}^{t} C(s)\, dx(s) \triangleq \int_{t_0}^{t} C(s)A(s)x(s)\, ds + \int_{t_0}^{t} C(s)M(s)\, dw(s). \tag{107}$$

Thus, we are now in a position to study the filtering problem.

6. THE WIENER-HOPF EQUATION

We suppose that the "signal" $x(t)$ is generated by the stochastic linear differential equation

$$dx = A(t)x\,dt + B(t)q(t)\,dw, \qquad x(t_0) = x_0, \qquad E\{x_0\} = 0 \quad (108)^{16}$$

and that the "observation" $z(t)$ is generated by the stochastic linear differential equation

$$dz = C(t)x\,dt + r(t)\,dw_1, \qquad z(t_0) = z_0, \tag{109}$$

where $R(t) = r(t)W_1r^*(t)$ is positive-definite, cov $[w(t), w_1(\tau)] = \mathbf{0}$ and $E\{\langle \mathrm{w}(t), w_1(\tau)\rangle\} = 0$ for all t and τ. Our filtering problem can now be stated as follows:

FILTERING PROBLEM: *Given $z(s)$ for $t_0 \leqq s \leqq t$, determine an estimate $\hat{x}(t_1|\,t)$ of $x(t_1)$ of the form*

$$\hat{x}(t_1 \mid t) = \int_{t_0}^{t} \mathbf{A}(t_1, s)\,dz(s) = \int_{t_0}^{t} \mathbf{A}(t_1, s)\,C(s)x(s)\,ds + \int_{t_0}^{t} \mathbf{A}(t_1, s)r(s)\,dw_1(s) \tag{110}$$

[*where the bounded linear-transformation-valued function $\mathbf{A}(\cdot,\cdot)$ is regulated in both arguments*], *with the property that*

$$E\{\langle x^*, x(t_1) - \hat{x}(t_1|\,t)\rangle^2\}, \qquad x^* \in H^* = H \tag{111}$$

is minimized for all x^* (i.e., the expected squared error in estimating any continuous linear functional of the signal is minimized).

The following theorem, which involves a "Wiener–Hopf equation," gives the basic necessary and sufficient condition for $\hat{x}(t_1|\,t)$ to be a solution of the filtering problem.

THEOREM 6.1 (Wiener–Hopf Equation). *Let $\tilde{x}(t_1|\,t) = x(t_1) - \hat{x}(t_1|\,t)$. Then $\hat{x}(t_1|\,t)$ is a solution of the filtering problem if and only if*

$$E\{\tilde{x}(t_1|\,t) \circ [z(\sigma) - z(\tau)]\} = \mathbf{0} \tag{112}$$

for all σ, τ with $t_0 \leqq \tau < \sigma < t$, or equivalently, if and only if

$$\operatorname{cov}[x(t_1), z(\sigma) - z(\tau)] - \operatorname{cov}\left[\int_{t_0}^{t} \mathbf{A}(t_1, s)\,dz(s), z(\sigma) - z(\tau)\right] = \mathbf{0} \tag{113}$$

for all σ, τ with $t_0 \leqq \tau < \sigma < t$.

Proof. Since $E\{x_0\} = 0$, $E\{z(\sigma) - z(\tau)\} = 0$ and so it is clear that (113) and (112) are equivalent.

[16] This restriction is inessential for if $E\{x_0\} \neq 0$, then we would replace $x(t)$ by $x(t) - \Phi(t, t_0)E\{x_0\}$.

Now let x^* be a fixed element of $H^* = H$ and let $X(x^*)$ be the space of all real random variables of the form $\langle x^*, x(\cdot)\rangle$ where $x(\cdot) \in L^2(\Omega, H)$. Define an inner product on $X(x^*)$ by $E\{\langle x^*, x(\cdot)\rangle \cdot \langle x^*, y(\cdot)\rangle\}$ and let $U(x^*)$ denote the subspace of $X(x^*)$ generated by elements of the form

$$\langle x^*, y(a)\rangle = \left\langle x^*, \int_{t_0}^{a} \mathbf{B}(t_1, s)\, dz(s)\right\rangle, a \leqq t, \tag{114}$$

where $\mathbf{B}(t_1, s)$ is regulated.[17] By the well-known orthogonal projection lemma (see, for example, Dunford and Schwartz, 1958), $\hat{x}(t_1|\, t)$ will be a solution of the filtering problem if and only if $\tilde{x}(t_1|\, t)$ is orthogonal to $U(x^*)$ in $X(x^*)$ for every x^*. In other words, $\hat{x}(t_1|\, t)$ is a solution of the filtering problem if and only if

$$E\{\langle x^*, \tilde{x}(t_1|\, t)\rangle \langle x^*, y(a)\rangle\} = 0 \tag{115}$$

for all x^* where

$$y(a) = \int_{t_0}^{a} \mathbf{B}(t_1, s)\, dz(s), a \leqq t \tag{116}$$

and $\mathbf{B}(t_1, s)$ is regulated.

So let us first suppose that (112) holds. We observe that, in view of Corollary 2.3,

$$E\{\langle x^*, \tilde{x}(t_1|\, t)\rangle \langle x^*, y(a)\rangle\} = x^* E\{\tilde{x}(t_1|\, t) \circ y(a)\}\, x. \tag{117}$$

However, $y(a)$ is given by (114); thus, if $\mathbf{B}(t_1, s)$ is a step function

$$\begin{aligned} E\{\tilde{x}(t_1|\, t) \circ y(a)\} &= \sum E\{\tilde{x}(t_1|\, t) \circ \mathbf{B}_j[z(\sigma_j) - z(\tau_j)]\} \\ &= \sum E\{\tilde{x}(t_1|\, t_0) \circ [z(\sigma_j) - z(\tau_j)]\}\mathbf{B}_j^* = \mathbf{0} \end{aligned} \tag{118}$$

since (112) is satisfied. It follows by "an approximation argument" (or say from Holder's inequality) that $E\{\tilde{x}(t_1|\, t) \circ y(a)\} = \mathbf{0}$ for any $y(a)$ given by (114) and hence, that

$$x^* E\{\tilde{x}(t_1|\, t) \circ y(a)\} x = 0 \tag{119}$$

for all x^*. Therefore, $\hat{x}(t_1|\, t)$ is a solution of the filtering problem.

On the other hand, let us suppose that $\hat{x}(t_1|\, t)$ is a solution of the filtering problem. If we assume that (112) does *not* hold, then

$$E\{\tilde{x}(t_1|\, t) \circ [z(\sigma) - z(\tau)]\} = \operatorname{cov}[\tilde{x}(t_1|\, t), z(\sigma) - z(\tau)] \neq \mathbf{0} \tag{120}$$

for some σ, τ with $t_0 \leqq \tau < \sigma \leqq t$. If we let $\mathbf{B}(t_1, s)$ be given by

[17] Note that $U(x^*)$ is the subspace of $X(x^*)$ generated by elements of the form $< x^*, \int_{t_0}^{t_0} \mathbf{C}(t_1, s) dz(s) >$ with $\mathbf{C}(t_1, s)$ regulated since, $\mathbf{B}(t_1, s)$, $t_0 \leqq s \leqq q$, regulated implies that $\mathbf{C}(t_1, s) = \mathbf{B}(t, s) = \mathbf{0}$, $q < s \leqq t$, is also regulated.

$$\mathbf{B}(t_1, s) = \begin{cases} \mathbf{0}, & s < \tau, \\ \operatorname{cov}[\tilde{x}(t_1|\, t),\ z(\sigma) - z(\tau)], & \tau \leqq s \leqq \sigma, \\ \mathbf{0}, & s > \sigma, \end{cases} \tag{121}$$

then $\mathbf{B}(t_1, s)$ is regulated (in fact, a step function) and

$$y(t) = \int_{t_0}^{t} \mathbf{B}(t_1, s)\, dz(s) = \operatorname{cov}[\tilde{x}(t_1 \mid t), z(\sigma) - z(\tau)](z(\sigma) - z(\tau)). \tag{122}$$

Thus,

$$\begin{aligned} E\{\langle x^*, \tilde{x}(t_1|\, t)\rangle \langle x^*, y(t)\rangle\} &= x^* E\{\tilde{x}(t_1|\, t) \circ y(t)\} x \\ &= x^* E\{\tilde{x}(t_1|\, t) \circ \operatorname{cov}[\tilde{x}(t_1|\, t),\ z(\sigma) - z(\tau)](z(\sigma) - z(\tau))\} x \\ &= x^* \operatorname{cov}[\tilde{x}(t_1|\, t), z(\sigma) - z(\tau)] \operatorname{cov}[\tilde{x}(t_1|\, t), z(\sigma) - z(\tau)]^* x. \end{aligned} \tag{123}$$

But (120) implies that there is some x for which the right-hand side of (123) is *not* zero; this is a contradiction. The proof of the theorem is now complete.

We observe that if $r(s)$ is essentially bounded and if $x(s)$ is essentially bounded (almost everywhere with respect to ω), then the theorem applies to estimates of the form

$$\hat{x}(t_1 \mid t) = \int_{t_0}^{t} \mathbf{A}(t_1, s)\, dz(s), \tag{124}$$

where $\mathbf{A}(\cdot, \cdot)$ is in L^2 in the pair (t_1, s). In particular this is true in the finite-dimensional case.

We shall use Theorem 6.1 to obtain an equation for the optimal filter in the next section.

7. THE OPTIMAL FILTER

We now develop the equation governing the optimal filter using some properties of covariances We rely heavily on Kalman and Bucy (1960) and begin with some lemmas.

LEMMA 7.1. *Let $\Phi(s)$ and $\Psi(s)$ be elements of $L^2(T, \mathfrak{L}(H, H))$. Then*

$$\operatorname{cov}\left[\int_{t_0}^{t} \Phi(s)\, dw(s), \int_{t_0}^{t} \Psi(s)\, dw_1(s)\right] = \mathbf{0} \tag{125}$$

and

$$\operatorname{cov}\left[\int_{t_0}^{t} \Phi(s)\, ds(s), \int_{t_0}^{t} \Psi(s)\, dw(s)\right] = \int_{t_0}^{t} \Phi(s) W \Psi^*(s)\, ds. \tag{126}$$

Proof. In view of Lemma 4.1 and the fact that $E\{\int \Phi(s)\, dw(s)\} = 0$, we can see that (126) holds by "an approximation argument." As for

(125), we need only establish that

$$E\left\{\left[\int_{t_0}^{t} R(s)\, dw(s)\right] \circ \left[\int_{t_0}^{t} S(a)\, dw_1(a)\right]\right\} = \mathbf{0} \qquad (127)$$

for step functions $R(s)$ and $S(a)$ since the general result will then follow by "an approximation argument." By adding $\mathbf{0}$'s if necessary, we may suppose that there are $a_0 < a_1 < \cdots < a_n$ in T such that

$$R(a) = \begin{cases} \mathbf{0}, & a < a_0 \\ R_j, & a_{j-1} \leqq a < a_j \\ \mathbf{0}, & a \geqq a_n, \end{cases} \qquad S(a) = \begin{cases} \mathbf{0}, & a < a_0 \\ S_k, & a_{k-1} \leqq a < a_k \\ \mathbf{0}, & a \geqq a_n. \end{cases} \qquad (128)$$

It then follows that

$$\begin{aligned} &\left[\int_{t_0}^{t} R(a)\, dw(a)\right] \circ \left[\int_{t_0}^{t} S(a)\, dw_1(a)\right] \\ &\qquad = \sum_{j=1}^{n} \sum_{k=1}^{n} R_j[w(a_j) - w(a_{j-1})] \circ [w_1(a_k) - w_1(a_{k-1})]\, S_k^* \end{aligned} \qquad (129)$$

and hence that (127) holds since cov $[w(t), w_1(\tau)] = \mathbf{0}$ for all t, τ.

Now, for simplicity of exposition, let us set $\Delta z(\sigma) = z(\sigma) - z(t_0)$ where $t_0 \leqq \sigma \leqq t$. Then

$$\begin{aligned} \Delta z(\sigma) &= \int_{t_0}^{\sigma} C(s)x(s)\, ds + \int_{t_0}^{\sigma} r(s)\, dw_1(s) \\ &= \int_{t_0}^{\sigma} C(s)\Phi(s, t_0)x_0\, ds + \int_{t_0}^{\sigma} C(s) \\ &\qquad \cdot \left[\int_{t_0}^{s} \Phi(s, a)B(a)q(a)\, dw(a)\right] ds + \int_{t_0}^{\sigma} r(s)\, dw_1(s), \end{aligned} \qquad (130)$$

where $\Phi(\cdot, \cdot)$ is the fundamental linear transformation of the system $\dot{h} = A(t)h$. Since $C(s)$ and $\Phi(s, a)$ are regulated and $B(a)q(a)$ is in $L^2(T, \mathfrak{L}(H, H))$, we deduce from Corollary 4.7 that

$$\begin{aligned} \Delta z(\sigma) &= \int_{t_0}^{\sigma} C(s)\Phi(s, t_0)x_0\, ds \\ &\qquad + \int_{t_0}^{\sigma} \left[\int_{a}^{\sigma} C(s)\Phi(s, a)\, ds\right] B(a)q(a)\, dw(a) \\ &\qquad\qquad + \int_{t_0}^{\sigma} r(s)\, dw_1(s). \end{aligned} \qquad (131)$$

To simplify notation, we shall let $\varphi_\sigma(a)$ be given by

$$\varphi_\sigma(a) = \int_a^\sigma C(s)\Phi(s, a)\, ds \tag{132}$$

We then have:

LEMMA 7.2. *Suppose that* $\operatorname{cov}[w(t), x_0] = \mathbf{0}$ *for all* t *and that* $K(t, s)$ *is in* $L^2(T \times T, \mathcal{L}(H, H))$, *Then*

$$\operatorname{cov}\left[\int_{t_0}^t K(t, s)\, dw(s), \Delta z(\sigma)\right] = \int_{t_0}^\sigma K(t, s) W q^*(s) B^*(s) \varphi_\sigma^*(s)\, ds. \tag{133}$$

Proof. Setting $\Psi(a) = \varphi_\sigma(a)B(a)q(a)$ for $a \leqq \sigma$ and $\Psi(a) = \mathbf{0}$ for $a > \sigma$ and noting that $\operatorname{cov}[w(t), x_0] = \mathbf{0}$ implies that $\operatorname{cov}[\int_{t_0}^t K(t, s)\, dw(s), x_0] = \mathbf{0}$, we deduce from (131) and Lemma 7.1 that

$$\begin{aligned}\operatorname{cov}&\left[\int_{t_0}^t K(t, s)\, dw(s), \Delta z(\sigma)\right] \\ &= \operatorname{cov}\left[\int_{t_0}^t K(t, s)\, dw(s), \int_{t_0}^t \Psi(s)\, dw(s)\right] \\ &= \int_{t_0}^t K(t, s) W \Psi^*(s)\, ds.\end{aligned} \tag{134}$$

The lemma follows immediately.

Now if $a(t)$ and $b(t)$ are random processes with $\operatorname{cov}[a(t), b(t)] = h(t)$ (a "sure" function), then it is natural to set

$$\frac{d}{dt}\operatorname{cov}[a(t), b(t)] = \frac{d}{dt} h(t) = \dot{h}(t) \tag{135}$$

whenever $\dot{h}(t)$ exists. Bearing this in mind, we have

COROLLARY 7.3. *If* $\operatorname{cov}[w(t), x_0] = \mathbf{0}$, *if* $K(t, s)$ *is in* $L^2(T \times T, \mathcal{L}(H, H))$ *and if* $\partial K(t, s)/\partial t$ *exists, is regulated in* t *and is* L^2 *in* s, *then, for* $\sigma < t$,

$$\frac{d}{dt}\operatorname{cov}\left[\int_{t_0}^t K(t, s)\, dw(s), \Delta z(\sigma)\right] = \operatorname{cov}\left[\int_{t_0}^t \frac{\partial K(t, s)}{\partial t}\, dw(s), \Delta z(\sigma)\right]. \tag{136}$$

COROLLARY 7.4. *If* $\mathrm{cov}[w_1(t), x_0] = \mathbf{0}$, *then, for* $\sigma < t$,

$$\frac{d}{dt}\,\mathrm{cov}\,[x(t), \Delta z(\sigma)] = \mathrm{cov}\,[A(t)x(t), \Delta z(\sigma)]. \tag{137}$$

Proof. Let $K(t, s) = \Phi(t, s)B(s)q(s)$ and apply Corollary 7.3.

We shall suppose from now on in this section that both $\mathrm{cov}[w(t), x_0] = \mathbf{0}$ and $\mathrm{cov}[w_1(t), x_0] = \mathbf{0}$ for all t. We then have:

LEMMA 7.5. *Suppose that* $L(t, s)$ *is regulated. Then*

$$\begin{aligned}\mathrm{cov}\left[\int_{t_0}^{t} L(t, s)\, dz(s), \Delta z(\sigma)\right] &= \int_{t_0}^{\sigma} \psi(t, s)B(s)q(s)Wq^*(s)B^*(s)\varphi_\sigma^*(s)\, ds \\ &\quad + \int_{t_0}^{\sigma} L(t, s)r(s)W_1\, r^*(s)\, ds \\ &\quad + \int_{t_0}^{t} L(t, s)C(s)\Phi(s, t_0)\, ds\cdot \mathbf{P}_0\cdot\varphi_\sigma^*(t_0),\end{aligned} \tag{138}$$

where

$$\psi(t, s) = \int_{s}^{t} L(t, a)C(a)\Phi(a, s)\, da, \tag{139}$$

$$\mathbf{P}_0 = \mathrm{cov}\,[x_0, x_0], \tag{140}$$

and $\varphi_\sigma(a)$ *is given by* (132).

Proof. First of all, we note that

$$\begin{aligned}\int_{t_0}^{t} L(t, s)\, dz(s) &= \int_{t_0}^{t} L(t, s)C(s)\left\{\Phi(s, t_0)\, x_0 \right.\\ &\quad \left. + \int_{t_0}^{s} \Phi(s, a)B(a)q(a)\, dw(a)\right\} ds \\ &\quad + \int_{t_0}^{t} L(t, s)r(s)\, dw_1(s).\end{aligned} \tag{141}$$

However, in view of (130), Lemma 7.1, the "independence" of x_0 and both $w(t)$ and $w_1(t)$, and Corollary 4.7 (Fubini), we readily see that

$$
\begin{aligned}
\operatorname{cov}&\left[\int_{t_0}^{t} L(t, s)\, dz(s), \Delta z(\sigma)\right] \\
&= \operatorname{cov}\left[\int_{t_0}^{t} \psi(t, s)B(s)q(s)\, dw(s), \Delta z(\sigma)\right] \\
&\quad + \int_{t_0}^{t} L(t, s)r(s)W_1 r^*(s)\, ds \\
&\quad + \int_{t_0}^{t} L(t, s)C(s)\Phi(s, t_0)\, ds\, \mathbf{P}_0\, \varphi_\sigma{}^*(t_0).
\end{aligned} \tag{142}
$$

The lemma then follows immediately from Lemma 7.2.

COROLLARY 7.6 *If $\partial L(t, s)/\partial t$ exists and is regulated, then, for $\sigma < t$,*

$$
\begin{aligned}
\frac{d}{dt}&\operatorname{cov}\left[\int_{t_0}^{t} L(t, s)\, dz(s), \Delta z(\sigma)\right] \\
&= \operatorname{cov}\left[\int_{t_0}^{t} \frac{\partial L(t, s)}{\partial t}\, dz(s) + L(t, t)C(t)x(t), \Delta z(\sigma)\right].
\end{aligned} \tag{143}
$$

(The proof of the corollary is a straightforward calculation and is therefore omitted.)

These lemmas and corollaries lead to the following theorem:

THEOREM 7.7. *Suppose that there is a solution of the filtering problem of the form*

$$\hat{x}(t \mid t) = \int_{t_0}^{t} L(t, s)\, dz(s) \tag{144}$$

with $\partial L(t, s)/\partial t$ regulated. Then

$$\frac{\partial L(t, s)}{\partial t} = A(t)L(t, s) - L(t, t)C(t)L(t, s) \tag{145}$$

for $t_0 \leqq s \leqq t$.

Proof. Since $\hat{x}(t \mid t)$ is a solution of the filtering problem, we have, by virtue of Theorem 6.1,

$$\frac{d}{dt}\operatorname{cov}[x(t), \Delta z(\sigma)] = \frac{d}{dt}\operatorname{cov}\left[\int_{t_0}^{t} L(t, s)\, dz(s), \Delta z(\sigma)\right]. \tag{146}$$

It follows from Corollaries 7.4, 7.6, and Theorem 6.1 that

$$
\begin{aligned}
\operatorname{cov}&\left[\int_{t_0}^{t} \left\{A(t)L(t, s) - \frac{\partial L(t, s)}{\partial t}\right.\right. \\
&\qquad \left.\left. - L(t, t)C(t)L(t, s)\right\} dz(s), \Delta z(\sigma)\right] = \mathbf{0}
\end{aligned} \tag{147}
$$

and hence that

$$\operatorname{cov}\left[\int_{t_0}^{t}\left\{A(t)L(t,s) - \frac{\partial L(t,s)}{\partial t} - L(t,t)C(t)L(t,s)\right\} dz(s), z(\sigma) - z(\tau)\right] = \mathbf{0} \tag{148}$$

since $z(\sigma) - z(\tau) = \Delta z(\sigma) - \Delta z(\tau)$. Setting

$$\Delta(t,s) = A(t)L(t,s) - \frac{\partial L(t,s)}{\partial t} - L(t,t)C(t)L(t,s), \tag{149}$$

we observe that (148) implies that

$$\hat{y}(t\,|\,t) = \int_{t_0}^{t} [L(t,s) + \Delta(t,s)]\, ds \tag{150}$$

satisfies (113) and hence is a solution of the filtering problem. As a consequence of the orthogonal projection lemma (see Dunford and Schwartz, 1958),

$$E\{\langle x^*, \hat{y}(t\,|\,t) - \hat{x}(t\,|\,t)\rangle^2\} = 0 \tag{151}$$

for all $x^* \in H$. In other words,

$$x^* \operatorname{cov}\left[\int_{t_0}^{t} \Delta(t,s)\, dz(s), \int_{t_0}^{t} \Delta(t,s)\, dz(s)\right] x = 0 \tag{152}$$

for all x. But

$$\operatorname{cov}\left[\int_{t_0}^{t} \Delta(t,s)\, dz(s), \int_{t_0}^{t} \Delta(t,s)\, dz(s)\right] = \int_{t_0}^{t} \Delta(t,s)R(s)\Delta^*(t,s)\, ds + \mathbf{Q}, \tag{153}$$

where $\mathbf{Q}$ is positive. Since $R(s)$ is *positive-definite* for all s, we immediately conclude that $\Delta(t,s) = \mathbf{0}$.

For simplicity of exposition we shall drop the "$|\,t$" from now on. Thus, we have:

COROLLARY 7.8. *Under the hypotheses of Theorem 7.7, $\hat{x}(t)$ satisfies the linear stochastic differential equation*

$$d\hat{x} = [A(t) - K(t)C(t)]\hat{x}\, dt + K(t)C(t)x\, dt + K(t)r(t)\, dw_1 \tag{154}$$

with $\hat{x}(t_0) = 0$ and where $K(t) = L(t,t)$.

Proof. We know that

$$\hat{x}(t) = \int_{t_0}^{t} L(t, s)\, dz(s) = \int_{t_0}^{t} L(t, s)C(s)x(s)\, ds + \int_{t_0}^{t} L(t, s)r(s)\, dw_1(s) \tag{155}$$

and hence that

$$\int_{t_0}^{t} [A(s) - K(s)C(s)]\hat{x}(s)\, ds = \int_{t_0}^{t} \Gamma(s) \left[\int_{t_0}^{s} L(s, a)C(a)x(a)\, da \right] ds + \int_{t_0}^{t} \Gamma(s) \left[\int_{t_0}^{s} L(s, a)r(a)\, dw_1(a) \right] ds \tag{156}$$

where $\Gamma(s) = A(s) - K(s)C(s)$. Applying Corollary 4.7 and the standard Fubini theorem, we have

$$\int_{t_0}^{t} \Gamma(s) \left[\int_{t_0}^{s} L(s, a)C(a)x(a)\, da \right] ds = \int_{t_0}^{t} \left[\int_{a}^{t} \Gamma(s)L(s, a)\, ds \right] C(a)x(a)\, da, \tag{157}$$

$$\int_{t_0}^{t} \Gamma(s) \left[\int_{t_0}^{s} L(s, a)r(a)\, dw_1(a) \right] ds = \int_{t_0}^{t} \left[\int_{a}^{t} \Gamma(s)L(s, a)\, ds \right] r(a)\, dw_1(a). \tag{158}$$

It then follows from (145) that

$$\begin{aligned} \int_{t_0}^{t} \Gamma(s)\hat{x}(s)\, ds &= \int_{t_0}^{t} \left[\int_{a}^{t} \frac{\partial L}{\partial s}(s, a)\, ds \right] dz(a) \\ &= \int_{t_0}^{t} \{L(t, a) - L(a, a)\}\, dz(a) \\ &= \hat{x}(t) - \int_{t_0}^{t} K(s)\, dz(a) \end{aligned} \tag{159}$$

and thus the corollary is established.

COROLLARY 7.9. *Under the hypotheses of Theorem 7.7, $\tilde{x}(t)$ satisfies the* *linear stochastic differential equation*

$$d\tilde{x} = [A(t) - K(t)C(t)]\,\tilde{x}\,dt + B(t)q(t)\,dw - K(t)r(t)\,dw_1 \tag{160}$$

with $\tilde{x}(t_0) = x_0$.

Corollaries 7.8 and 7.9 are at the heart of the development of Kalman and Bucy (1960). Continuing in the same vein, we have the following theorem:

THEOREM 7.10. *Suppose that the conditions of Theorem* 7.7 *are satisfied. Then*

$$K(t) = P(t)C^*(t)R^{-1}(t), \tag{161}$$

where $P(t)$ *is a solution of the Riccati-type equation*

$$\dot{P} = AP + PA^* - PC^*R^{-1}CP + Q \tag{162}$$

with $P(t_0) = P_0 = \operatorname{cov}[x_0, x_0]$ *and* $Q = BqWq^*B^*$.

Proof. Let us set

$$y(\sigma) = \int_{t_0}^{\sigma} C(s)x(s)\,ds = \Delta z(\sigma) - \int_{t_0}^{\sigma} r(s)\,dw_1(s). \tag{163}$$

Then, we have, by direct computation,

$$\begin{aligned}\frac{d}{d\sigma}\operatorname{cov}[\tilde{x}(t), y(\sigma)] &= \Phi(t, t_0)P_0\Phi^*(\sigma, t_0)C^*(\sigma)\\ &\quad+ \int_{t_0}^{\sigma} \Phi(t, s)Q(s)\Phi^*(t_0, s)\,ds\Phi^*(\sigma, t_0)C^*(\sigma)\\ &\quad- \psi(t, t_0)P_0\Phi^*(\sigma, t_0)C^*(\sigma)\\ &\quad- \int_{t_0}^{\sigma} \psi(t, s)Q(s)\Phi^*(t_0, s)\,ds\Phi^*(\sigma, t_0)C^*(\sigma)\end{aligned} \tag{164}$$

and, by Theorem 6.1

$$\operatorname{cov}[\tilde{x}(t), y(\sigma)] = \int_{t_0}^{\sigma} L(t, s)R(s)\,ds, \tag{165}$$

where $\psi(t, s)$ is given by (139). It follows that for $\sigma < t$

$$\begin{aligned}L(t, \sigma)R(\sigma) = \Big[&\{\Phi(t, t_0) - \psi(t, t_0)\}P_0\\ &+ \int_{t_0}^{\sigma} \{\Phi(t, s) - \psi(t, s)\}Q(s)\Phi^*(t_0, s)\,ds\Big]\Phi^*(\sigma, t_0)C^*(\sigma).\end{aligned} \tag{166}$$

Since any regulated function is equivalent (in an almost everywhere sense) to a function continuous on the left, we can take limits as σ approaches t from below in (166) and thus deduce that $K(t) = P(t)C^*(t)R^{-1}(t)$ where

$$P(t) = \left[\{\Phi(t, t_0) - \psi(t, t_0)\}P_0 + \int_{t_0}^{t} \{\Phi(t, s) - \psi(t, s)\}Q(s)\Phi^*(t_0, s)\, ds \right] \Phi^*(t, t_0). \tag{167}$$

Clearly $P(t_0) = P_0$.

Now we note that

$$\frac{\partial\Phi(t, s)}{\partial t} = A(t)\Phi(t, s), \qquad \frac{\partial\Phi^*(t, t_0)}{\partial t} = \Phi^*(t, t_0)A^*(t) \tag{168}$$

and that

$$\frac{\partial\psi(t, s)}{\partial t} = L(t, t)C(t)\Phi(t, s) + \int_{s}^{t} \{A(t)L(t, a) - L(t, t)C(t)L(t, a)\}C(a)\Phi(a, s)\, da \tag{169}$$

by Theorem 7.7. Moreover, $L(t, t)C(t) = P(t)C^*(t)R^{-1}(t)C(t)$. The theorem then follows from these relations by direct differentiation of (167).

Corollary 7.11. *$P(t)$ is symmetric.*

Proof. If $P(t)$ satisfies (162), then $P^*(t)$ also satisfies (162). As $P_0 = P_0^*$, the result follows from the uniqueness of solutions of Banach-space differential equations (Dieudonne, 1960).

Corollary 7.12. $P(t) = \operatorname{cov}[\tilde{x}(t), \tilde{x}(t)]$.

Proof. We have

$$\operatorname{cov}[\tilde{x}(t), \tilde{x}(t)] = \operatorname{cov}[\tilde{x}(t), x(t)] - \operatorname{cov}[\tilde{x}(t), \hat{x}(t)], \tag{170}$$

$$\operatorname{cov}[\tilde{x}(t), x(t)] = \operatorname{cov}[x(t), x(t)] - \operatorname{cov}[\hat{x}(t), x(t)], \tag{171}$$

$$\operatorname{cov}[x(t), x(t)] = \left[\Phi(t, t_0)P_0 + \int_{t_0}^{t} \Phi(t, s)Q(s)\Phi^*(t_0, s)\, ds\right]\Phi^*(t, t_0), \tag{172}$$

$$\operatorname{cov}[\hat{x}(t), x(t)] = \left[\psi(t, t_0)P_0 + \int_{t_0}^{t} \psi(t, s)Q(s)\Phi^*(t_0, s)\, ds\right]\Phi^*(t, t_0), \tag{173}$$

and, by the lemma which follows the corollary,

$$\mathrm{cov}[\tilde{x}(t), \ \hat{x}(t)] = \mathbf{0}. \tag{174}$$

Thus, the corollary is established.

LEMMA 7.13.[18] $\mathrm{cov}[\tilde{x}(t), \hat{x}(t)] = \mathbf{0}$.

Proof. We have

$$\mathrm{cov}\,[\tilde{x}(t), \hat{x}(t)] = \mathrm{cov}\,[x(t), \hat{x}(t)] - \mathrm{cov}\,[\hat{x}(t), \hat{x}(t)], \tag{175}$$

$$\mathrm{cov}\,[x(t), \hat{x}(t)] = \int_{t_0}^{t} \Phi(t, s)Q(s)\psi^*(t, s)\,ds + \Phi(t, t_0)P_0\psi(t, t_0)^*, \tag{176}$$

$$\begin{aligned}\mathrm{cov}\,[\hat{x}(t), \hat{x}(t)] = &\int_{t_0}^{t} L(t, s)R(s)L^*(t, s)\,ds \\ &+ \int_{t_0}^{t} \psi(t, s)Q(s)\psi^*(t, s)\,ds + \psi(t, t_0)P_0\psi^*(t, t_0).\end{aligned} \tag{177}$$

However, $L(t, s)R(s)$ is given by (166) and so the lemma follows by a straightforward calculation and the standard Fubini theorem.

We note that the operator Riccati equation (162) is discussed in Kalman *et al.* (1967) and in Falb and Kalman (1966). In particular, it is shown that (162) has a unique solution which is defined on the entire interval of definition of A, C, R and Q in these references. Let us denote this solution by $P_1(t)$ and let us set $K_1(t) = P_1(t)C^*(t)R^{-1}(t)$ Then the linear stochastic differential equation

$$\begin{aligned}d\hat{x}_1 = &[A(t) - K_1(t)C(t)]\hat{x}_1\,dt \\ &+ K_1(t)C(t)x\,dt + K_1(t)r(t)\,dw_1\end{aligned} \tag{178}$$

has a solution $\hat{x}_1(t)$ with $\hat{x}_1(t_0) = 0$. If $\Psi(t, s)$ is the fundamental linear transformation of the linear differential equation

$$\dot{h} = [A(t) - K_1(t)C(t)]h, \tag{179}$$

then

$$\hat{x}_1(t) = \int_{t_0}^{t} L_1(t, s)\,dz(s), \tag{180}$$

where

$$L_1(t, s) = \Psi(t, s)K_1(s). \tag{181}$$

We observe that $L_1(t, t) = K_1(t)$ and that

[18] Lemma 7.13 can also be proven using Theorem 6.1 and "an approximation argument."

$$\frac{\partial L_1(t, s)}{\partial t} = A(t)L_1(t, s) - L_1(t, t)C(t)L_1(t, s) \tag{182}$$

We now have:

THEOREM 7.14. *$\hat{x}_1(t)$ is a solution of the filtering problem.*

Proof. Let us set $\tilde{x}_1(t) = x(t) - \hat{x}_1(t)$. Then, by theorem 6.1, we need only show that $\text{cov}[\tilde{x}_1(t), \Delta z(\sigma)] = \mathbf{0}$ for $\sigma < t$. If $\psi_1(t, s)$ is given by

$$\psi_1(t, s) = \int_s^t L_1(t, a)C(a)\Phi(a, s)\, da, \tag{183}$$

then it will follow from (168) and (182), by direct differentiation that

$$P_1(t) = \left[\{\Phi(t, t_0) - \psi_1(t, t_0)\}P_0 + \int_{t_0}^t \{\Phi(t, s) - \psi_1(t, s)\}Q(s)\Phi^*(t_0, s)\, ds \right] \Phi^*(t, t_0). \tag{184}$$

Now, a direct computation shows that

$$\begin{aligned} \frac{d}{d\sigma} \text{cov}\, [\tilde{x}_1(t), y(\sigma)] = {} & \Phi(t, t_0)P_0\Phi^*(\sigma, t_0)C^*(\sigma) \\ & + \int_{t_0}^{\sigma} \Phi(t, s)Q(s)\Phi^*(t_0, s)\, ds\Phi^*(\sigma, t_0)C^*(\sigma) \\ & - \psi_1(t, t_0)P_0\Phi^*(\sigma, t_0)C^*(\sigma) \\ & - \int_{t_0}^{\sigma} \psi_1(t, s)Q(s)\Phi^*(t_0, s)\, ds\Phi^*(\sigma, t_0)C^*(\sigma) \end{aligned} \tag{185}$$

and that

$$\frac{d}{d\sigma} \text{cov} \left[\tilde{x}_1(t), \int_{t_0}^{\sigma} r(s)\, dw_1(s) \right] = -L_1(t, \sigma)R(\sigma). \tag{186}$$

Letting $\phi(t, \sigma) = d\{\text{cov}[\tilde{x}_1(t), \Delta z(\sigma)]\}/d\sigma$, we have $\phi(\sigma, \sigma) = \mathbf{0}$ and, after some straightforward computations,

$$\frac{\partial \phi(t, \sigma)}{\partial t} = \{A(t) - K_1(t)C(t)\}\phi(t, \sigma) \tag{187}$$

It follows that

$$\phi(t, \sigma) = \Psi(t, \sigma)\phi(\sigma, \sigma) = \mathbf{0} \tag{188}$$

and hence that $\text{cov}[\tilde{x}_1(t), \Delta z(\sigma)] = \text{cov}[\tilde{x}_1(t), \Delta z(t_0)] = \mathbf{0}$. Thus, the theorem is established.

In essence, we have shown in Theorem 7.14 that (162) provides the basis for a complete solution to the filtering problem. Moreover, this implies that the duality theorem of Kalman and Bucy (1960) also holds in the infinite-dimensional case.

8. CONCLUDING REMARKS

We have developed the theory of the Kalman–Bucy filter in a Hilbert-space context. Our development depended upon our definition of the covariance as a bounded linear transformation, our introduction of the stochastic integral, our use of a Fubini-type theorem involving the interchange of stochastic and Lebesgue integration, and some calculations with covariances. We avoided several complex measure-theoretic problems relating to the notions of martingale and semimartingale for H-valued random processes since these ideas were not needed in our treatment; however, we plan to examine these questions in a subsequent paper. We also note that it is possible to discuss nonlinear filtering in the infinite-dimensional realm using the notions developed here and a suitable generalization of Ito's lemma. Finally, we note that we have obtained a fully rigorous theory for the finite-dimensional case and have not required either the delta function or Ito's lemma.

RECEIVED: October 10, 1966; revised January 4, 1967.

REFERENCES

BEUTLER, F. J., (1963), Multivariate wide-sense Markov processes and prediction theory. *Ann. Math. Stat.* **34,** 421–438.

DIEUDONNE, J., (1960), "Foundations of Modern Analysis" Academic Press, New York.

DOOB, J. L., (1953), "Stochastic Processes." Wiley, New York.

DUNFORD, N. AND J. T. SCHWARTZ, (1958), "Linear Operators, Part I: General Theory." Interscience, New York.

FALB, P. AND Kleinman, D. L. (1966), Remarks on the infinite dimensional Riccati equation. *IEEE Trans. Aut. Cont.* **AC-11,** 534–536.

KALMAN, R. E. AND BUCY, R. S. (1960), New results in linear filtering and prediction theory. *Trans. ASME*, *J. Basic Eng.* **83D,** 95–108.

KALMAN, R. E., FALB, P. L., AND ARBIB, M. A. (1967), "Topics in Mathematical System Theory." McGraw-Hill, New York.

KANTOROVICH, L. V. AND AKILOV, G. P. (1964), "Functional Analysis in Normed Spaces." Pergamon, New York.

KUSHNER, H. J. (1964), On the differential equations satisfied by conditional probability densities of Markov processes with applications. *Tech. Rept.*, Research Institute for Advanced Studies, Baltimore, Maryland.