In their seminal paper, Fujisaki, Kallianpur and Kunita [1] showed how the least squares estimate of a signal contained in additive white noise can be represented as a stochastic integral with respect to innovation process, the integral being adapted to the observation process. The difficulty with this representation is that in general this estimate is not useful for computing the estimate since the innovations process depends on the estimate of the signal itself. In this paper we discuss representation of the estimate directly in terms of the observation process. In doing so, we derive new results on...
multiple integral expansions for square-integrable functionals of the observation process and show the connection of this work to the theory of contraction operators on Fock space. This letter development is due to Nelson and Segal.

We also present several applications of these results to determining sub-optimal filters.
1. Introduction

In their seminal paper, Fujisaki, Kallianpur and Kunita [1] showed how the best least squares estimate of a signal contained in additive white noise can be represented as a stochastic integral with respect to an innovation process, the integral being adapted to the observation process. The difficulty with this representation is that in general this estimate is not useful for computing the estimate since the innovations process depends on the estimate of the signal itself. In this paper we discuss representation of the estimate directly in terms of the observation process. In doing so, we derive new results on multiple integral expansions for square-integrable functionals of the observation process and show the connection of this family of sub—algebras \( F_t \). We assume familiarity with the definition of conditional expectation, and thus they give a framework for considering best Proof. See Ito the natural concept of a polynomial in the process through a change of measure, the form therefore, an probability space. Wiener's homogeneous chaos expansion, which as an example of the general theory presented later, decomposes \( L^2([0,T]) \) into a direct sum of Hilbert space tensor products. Indeed if \( H \equiv \mathbb{R} \), \( H_n \equiv L^2([0,T]) \) is a simple application of \( (a_n) \) and \( \widetilde{a} \) demonstrates that \( H_n \) is a Hilbert space for every \( n \) and that \( H \equiv \mathbb{R} \) for \( n \mathbb{N} \) where \( l \) is defined in the sense of the inner product (x,y)\( _{E2} \). In fact we have more:

**Theorem 1** (Ito-Wiener):

\[ L^2([0,T]) = H \oplus H \]

That is, for \( L^2([0,T]) \) kernels \( k_n \) exist such that

\[ f \Rightarrow k_n \]

(5)


Theorem 1 suggests the following natural question.

Suppose \( \mathcal{L}^2([0,T]) \) and \( \mathcal{E}^2([0,T]) \). Is it then true that \( \mathcal{K} \) and \( \mathcal{L} \) are kernels \( k_n \) or \( \hat{k}_n \) for \( T \), and if so, what are the kernels \( k_n \)?

1. Approved for public release; distribution unlimited.
\[ \sum_{k=0}^{n} e^{-\frac{1}{2} \beta t} x_k \left( e^{\frac{1}{2} \beta t} (s_{m+1-k}) \right)^2 + \frac{1}{2} \beta^2 \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \left( e^{\frac{1}{2} \beta t} (s_{m+1-i}) \right) \left( e^{\frac{1}{2} \beta t} (s_{m+1-j}) \right) \]

**Definition 3** 1) \( F \) will denote the projection of \( L^2([0,T]^n) \) onto \( L^2([0,T]^m) \)

\[ \langle f, g \rangle = \int_{[0,T]^m} f(s_1, \ldots, s_m) g(s_1, \ldots, s_m) ds_1 \ldots ds_m \]

where \( S \) is permutation group on \( n \) letters.

11) Let \( 0 < m \leq n \) and \( f, g \in L^2([0,T]^m) \)

\[ \langle f, g \rangle = \int_{[0,T]^m} f(s_1, \ldots, s_m) g(s_1, \ldots, s_m) ds_1 \ldots ds_m \]

where \( h \) is a measure on \( [0,T]^m \).

**Proof** To understand the decomposition of theorem 1. In fact, if \( \{ e_n \}_{n=1}^\infty \) is a complete orthonormal basis of \( L^2([0,T]^m) \) and \( \{ f_n \}_{n=1}^\infty \) is spanned by \( \{ e_n \}_{n=1}^\infty \), then \( Ito \) has shown that \( h = \sum_{n=1}^\infty \) \( e_n \) denotes closure.

**Theorem 3** Let \( n > 1 \) and \( f, g \in L^2([0,T]^m) \). Then \( \langle f, g \rangle \) has consequences that relate directly to the theory of contractions on \( L^2 \) spaces of Hilbert space tensor products presented in a later section. The point is that the multiplication formula can be used to study the integrability of \( k \)th order moments of the integral \( \int_t^e \), and, indeed, a direct application of (6) using lemma 1 and a recursion argument yields:

\[ \langle f, g \rangle = \int_{[0,T]^m} f(s_1, \ldots, s_m) g(s_1, \ldots, s_m) ds_1 \ldots ds_m \]

and \( f = g \) for \( t \leq T \).

**Proof** Let \( n = m = 1 \) and \( f, g \) be functions in \( L^2([0,T]^n) \).

Then \( \langle f, g \rangle = \int_{[0,T]^n} f(s_1, s_2) g(s_1, s_2) ds_1 ds_2 \) for any \( n \geq k \).

**Theorem 4** Let \( f, g \in L^2([0,T]^n) \). Then \( \langle f, g \rangle = 0 \) if and only if \( \| f \|_2 = 0 \).

**Proof** By (6) \( \| f \|_2 = 0 \) if and only if \( \langle f, f \rangle = 0 \).
Theorem 5 Under the hypotheses of (2)
i) \( P_0 \) is a probability measure and \( P \) and \( P_0 \) are mutually absolutely continuous with \( \frac{dP}{dP_0} = L(t) \).
ii) Under \( P_0 \), \( y(t) \) is a Brownian motion independent of \( x(t) \).
iii) \( x(t) \) has the same distribution under \( P_0 \) as under \( P \).

\( k(t) = \frac{E(x(t)|L(t))}{E(L(t))} \)

The kernels \( k_j \) and \( l_j \) depend only on the distribution of \( x(t) \).

The condition \( E\{\exp[\int_0^t h^2(s)ds]\} < \infty \) places strong restrictions on the growth of the moments of \( h^2(s)ds \).

Proof By iterating (16):

\[ L(t) = \int_0^t \int_0^t h^2(s)ds L(r)dy(s) \]

Continuing such iteration ad infinitum we derive the formal expression

\[ L(t) = \int_0^t \int_0^t h^2(s)ds L(r)dy(s) \]

The third equality follows from Theorem 5 ii) and iii), and the fourth equality by definition. By a similar calculation,

\[ E(L(t)|L(t)) = \int_0^t E[L(t)|L(t)]dL(t) \]

The third equality follows from Theorem 5 ii) and iii), and the fourth equality by definition. By a similar calculation,

\[ E(L(t)|L(t)) = \int_0^t E[L(t)|L(t)]dL(t) \]

Now substitute (17) into the term \( E(L(t)|P_t) \). We get:

\[ E[L(t)|P_t] = \int_0^t E[L(t)|L(t)]dL(t) \]

where \( k_j \) and \( l_j \) are as above and the infinite series both converge in the \( L^2(P) \) norm.

Remarks
1. The kernels \( k_j \) and \( l_j \) depend only on the a priori distribution of \( x(t) \).
2. The condition \( E\{\exp[\int_0^t h^2(s)ds]\} < \infty \) in (6) places strong restrictions on the growth of the moments of \( h^2(s)ds \).
3. Applications

The explicit formulas (14) and (15) can be applied to the design of suboptimal filters in various ways. For example, one naive approach would be to truncate the numerator and denominator of the ratio at finite orders and use the result as an approximate filter. As noted in the remarks after theorem 6, the kernels of the expansions do not involve the observations $y(t)$ and so can be computed off line. Theoretically then, it is possible to construct the truncated filter. This design, however, is difficult to analyze and assess; a more interesting use of theorem 6 involves finding estimates that are multiple integral expansions of finite order.

**Definition 5.** a) An expression of the form

$$c(t) = e_0(t) + \int_0^T e_n(t) \langle a_n(t) \rangle$$

with $e_n(t) \in L^2([0,T])$ is called an nth order expansion of $y(t)$.

b) An nth order expression $a(t)$, satisfying

$$E[f(t)-c(t)]^2 \leq E[f(t)-c(t)]^2$$

for any other nth order c(t), is called the best nth order estimate of f(t).

The best nth order estimate will be denoted by $\hat{f}(t)$ (with $n$ understood), and its kernels by $a_0, a_1, \ldots$. Given an order $n$, how can we find $\hat{f}(t)$, that is how can we determine $a_0, a_1, \ldots$? As it turns out, we can apply the multiplication formula to the filter expansion to write integral equations for the kernels $a_n$. Begin by considering the product $f(t)E_0[L(t)|P_t]^\perp$ of the estimate with the denominator of (13). If

$$E[f(t)E_0[L(t)|P_t]^\perp] = \int_0^T \langle a_n(t) \rangle$$

then the expansion of $E_0[L(t)|P_t]^\perp$ in (14) applies, and

$$
\begin{align*}
E[f(t)E_0[L(t)|P_t]^\perp] &= \int_0^T \langle a_n(t) \rangle \frac{E[f(t)L(t)]^\perp}{E[f(t)L(t)]^\perp} \\
&= \int_0^T \langle a_n(t) \rangle \frac{E[f(t)L(t)]^\perp}{E[f(t)L(t)]^\perp}
\end{align*}
$$

The second equality in (25) uses the identities

$$E[f(t)c(t)]^2 = \int_0^T \langle e_n(t) \rangle$$

which are easily demonstrated. Now under $P_0$ $y(t)$ is a Brownian motion and integrals of different orders are orthogonal. Thus, using (20) and

$$E[f(t)E_0[L(t)|P_t]^\perp] = \int_0^T \langle a_n(t) \rangle$$

in (25),

$$E[f(t)-f(t)]c(t) = E[f(t)-c(t)]$$

An application of lemma 2 shows that the second and third terms of the r.h.s of (26) are zero for all c(t). Clearly, the first term can be zero for all nth order c(t) if $k_n = e_n$ for $0 \leq n \leq r$, and this completes the proof.

The equations (26) are actually integral equations for the kernels $a_n(t)$ of the best nth order estimate, since the $a_n(t), 0 \leq n \leq r$, are found from $a_0(t), 0 \leq n \leq r$, by the formula (8). To illustrate, if $r=1$, $\lambda_1(t) = E(h(t))$ and

$$E[f(t)h(s)] = a_0(t)E_0[L(t)|P_t]^\perp$$

Solving for $a_0(t)$,

$$a_0(t) = \frac{E[f(t)h(s)|P_t]^\perp}{E_0[L(t)|P_t]^\perp}$$

This is the familiar Wiener-Hopf equation for the best linear estimate. In the best quadratic (r=2) case, the equations become more complicated. They are, assuming $E(h(s)) = 0$, $E(f(t)) = 0$ for simplicity,

$$a_0(t) = \frac{E[f(t)h(s)]}{E_0[L(t)|P_t]^\perp}$$

$$a_1(t,s) = \frac{E[f(t)h(s)-f(t)h(s)]}{E_0[L(t)|P_t]^\perp}$$

This is the familiar Wiener-Hopf equation for the best quadratic estimate.
In (27), \( \text{cov} \{ X_1, ..., X_n \} = E(X_1, ..., X_n) = \cdots = E(X_{n-1}, ..., X_n) \).

(27) shows how the kernels of different orders are dependent on one another. Though not a standard integral equation, (27) may be reduced, by using the solution of a related linear estimation problem, to a single integral equation for \( \sigma \). For fixed \( t \) this equation is of Fredholm type for \( \sigma(X,t) \) and can be solved by standard methods. We shall not go into this theory here.

The multiplication formula can also be used to derive the Kalman filter. Consider the simple case

\[
dx(t) = db(t), \quad x(0) = 0
\]

\[
dy(t) = x(t)dt + dw(t)
\]

where \( b(\cdot) \) and \( w(\cdot) \) are independent Brownian motions.

Then we can show that the optimal filter is

\[
\hat{P}(t) = \gamma^2(\sigma(t,s)dy(s))
\]

where \( \sigma_2(t,s) = \sigma(t,s) - \sigma(0,s) \).

The proof is simply to show that \( \sigma(t,s) \) can be chosen so that

\[
\hat{P}(t) = \int_0^t \sigma(t,s)dy(s) - \sigma(0,s)dy(s)
\]

\[
\text{or} \quad \int_0^t \sigma(t,s)dy(s) = \int_0^t \sigma(0,s)dy(s) - \sigma(t,s)dy(s)
\]

By expanding the l.h.s of (29) using (8), and equating kernels of different orders we derive the infinite set of equations.

\[
J_n(t, s) = \sum_{j=1}^n \sigma_j(t,s)J_{n-j}(s)
\]

(31)

It can now be shown that if (31) is satisfied for \( J=1 \), it is satisfied for all \( J \geq 1 \), a result following Hida. We use this mapping to identify \( \sigma(H) \) and \( H \).

\[
\int_0^t \sigma(t,s)dy(s) = \int_0^t \sigma(0,s)dy(s) - \sigma(t,s)dy(s)
\]

(30)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(32)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(33)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(34)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(35)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(36)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(37)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(38)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(39)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(40)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(41)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(42)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(43)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(44)

\[
\text{Sym}(\sigma(t,s)dy(s)) = \int_0^t \sigma(t,s)dy(s)
\]

(45)
properties and the precise statement of this is an important theorem of Nelson. Before we discuss this result it is useful to recall that conditional expectations on $L^2(\Omega, \mathcal{A}, \mu)$ can be characterised as linear positivity preserving operators which are idempotent, of norm $\leq 1$ and preserve constants. We also know that for $p \{1, \infty\}$, $p \neq 2$, all linear operators $T$ on $L^p(\Omega, \mathcal{A}, \mu)$, which are idempotent, contracting and such that $T1=1$ is necessarily a conditional expectation.

Theorem 3.1 (Nelson Hypercontractivity Theorem). Let $A: H \rightarrow K$ be a contraction. Then $I(A)$ is a contraction from $L^q(H) \rightarrow L^p(K)$ for $1 < q < \infty$ provided that

$$
||x||^2 \leq \frac{(p-1)}{p-1} \frac{1}{2} (40)
$$

If (40) does not hold then $I(A)$ is not a bounded operator from $L^q(H) \rightarrow L^p(K)$.


Acknowledgment

This research has been supported by the Air Force Office of Scientific Research under Grant No. AFSOR-77-3288.