MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING

S. K. Mitter and D. Ocone
Massachusetts Institute of Technology
Cambridge, MA

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 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.

2. Multiple Integrals and Filtering

In this section, we shall discuss applications of multiple integral expansions to the general filtering problem. We will consider the 'canonical' scalar filtering model:

$$y_t=\int_0^t h(x_s)ds + w_t \tag{1}$$

under the assumptions

a) x_t and w_t are independent processes

b) for some $T>0$ $E\int_0^T h^2(x_s)ds<\infty$ (2)

c) w_t is a standard Brownian motion

If $f_t(x_{(\cdot)})=f_t(x_s,s\leq t)$ is a causal functional of the signal x_t and $F_t^y\equiv\sigma\{y_s|0\leq s\leq t\}\equiv$ sub-σ-algebra generated by $y_s, 0\leq s\leq t$, then we are interested in calculating the optimal least squares estimate of $f_t(x_{(\cdot)})$

$E\{f_t(x_{(\cdot)})|F_t^y\}$ for $t\leq T$.

<u>Definition 1</u> y_t defined in (1) and (2) is called an <u>observation semi-martingale</u>.

Throughout, let (Ω,F,P) denote the underlying probability space.

Now $E\{f_t(x)|F_t^y\}\varepsilon L(\Omega,F_t^y,P)$ $(=\{F_t^y$ measurable rv's$\})$ by the definition of conditional expectation, and, therefore, any method that represents elements of $L(\Omega,F_t^y,P)$ in a simple and consistent way, say by expansion in terms of a simple class of functionals of $y_{(\cdot)}$ can be applied to the optimal estimate. In this work, we have adopted multiple integrals of the form $\int_0^t..\int_0^{s_{r-1}}k(t,s_1,..s_r)dy(s_r)..dy(s_1)$ as the basic objects of expansion. First, y_t is a stochastic translation of Brownian motion and through a change of measure, much Brownian theory can be carried over. Secondly, iterated integrals provide the natural concept of a polynomial in the y process and thus they give a framework for considering best quadratic, cubic, etc. suboptimal estimation procedures. Finally, when the kernel k of $z_t=\int_0^t..\int_0^{s_{r-1}}k dy_{s_r}..dy_s$, is separable, a construction of Brockett [2] realizes z_t recursively as the solution to a stochastic differential equation.

Accordingly, after developing some theory of multiple integral expansions we show how $E\{f_t(x_{(\cdot)})|F_t^y\}$ can be represented as a ratio of multiple integral expansions. The chief theoretical result about multiple integrals, the multiplication formula of theorem 2, is then used in conjunction with this representation to derive equations for the best suboptimal estimate of any order. The Kalman filter is derived and the quadratic filter discussed in detail as examples.

<u>Multiple Integrals</u>. In what follows, let $(b(t),F_t)$ denote a standard Brownian motion w.r.t. increasing family of sub-σ-algebras F_t. We assume familiarity with the stochastic integral $\int_0^t\phi_s db(s)$, where $\phi_s(u)$ is a measurable process adapted to $(F_t)_{t\geq 0}$.

<u>Definition 2</u> Let $f\varepsilon L^2(\widehat{[0,T]^n})\equiv\{f\varepsilon L^2([0,T]^n)|f \text{ symmetric}\}$. $I_t^{(m)}(f)$, the nth order multiple (or iterated) integral up to $t\leq T$ of f, is defined recursively by

$$I_t^{(n)}(f)=\int_0^t I_s^{(m-1)}f(s,..)db(s) \tag{3}$$

In (3), $f(s,..)$ is the function of $L^2(\widehat{[0,T]^{n-1}})$ formed by holding the first element of f fixed at s. Strictly speaking, for (3) to make sense it must be shown that $I_s^{(n-1)}(f(s,..))$ has a measurable version, but this can easily be done by approximating f with separable functions. Let $(f,g)=\int_0^t..\int_0^t f(s_1,..,s_n)g(s_1,..,s_n)ds_n..ds_1$ denote the inner product of $L^2([0,T]^n)$. By applying standard facts about stochastic integrals, the following basic properties of the multiple integral are derived: for any n and m, $t\leq T$, and $f\varepsilon L^2([0,T]^n)$, $g\varepsilon L^2([0,T]^m)$

a) $E\{I_t^{(n)}(f)\}=0$ (4)

b) $E\, I_t^{(n)}(f)I_t^{(m)}(g)=\begin{cases}0 & \text{if } m\neq n\\ 1/n!(f,g) & \text{if } m=n\end{cases}$

Note also that $I_t^{(n)}(f)$ depends only on the values of $f(s_1,..,s_n)$ for $s_1\geq s_2\geq..\geq s_n$. (3) adopts the useful convention of allowing f to be defined in all of $[0,T]^n$ by a symmetric extension.

Multiple integrals are useful in constructing Wiener's homogeneous chaos expansion, which as an example of the general theory presented later, decomposes $L^2(F_t^b)$ into a direct sum of Hilbert space tensor products. Indeed if $H_0\equiv R$, $H_n\equiv\{I_t^{(n)}(f)|f\varepsilon L^2([0,T]^n)\}$ $n\geq 1$ a simple application of (4) a) and b) demonstrates that H_n is a Hilbert space for every n and that $H_n\perp H_m$ for $n\neq m$ [where $\perp$ is defined in the sense of the inner product $(x,y)=Exy$]. In fact we have more:

<u>Theorem 1</u> (Ito-Wiener)

$$L^2(F_t^b)=H_0\oplus H_1\oplus H_2\oplus \ldots\ldots$$

That is, for $\phi\varepsilon L^2(F_t^b)$ kernels $\{k_n\}_{n=0}^{\infty}$ exist such that

$$\phi=k_0+\Sigma_{n=1}^{\infty}I_t^{(n)}(k_n) \tag{5}$$

Proof. See Ito [3] and Kallianpur [4].

Theorem 1 suggests the following natural question. Suppose $f\varepsilon L^2([0,T]^n)$ and $g\varepsilon L^2([0,T]^m)$. Is it then true that $I_t^{(n)}(f)I_t^{(m)}(g)\varepsilon L^2(F_t^b)$ for $t\leq T$, and if so, what are the kernels $\{k_i\}$, as in (5), such that $I_t^n(f)I_t^m(g)$

CH1486-0/79/0000-0329$00.75 © 1979 IEEE

$=k_o+\Sigma_{i=1}^{\infty} I_t^{(i)}(k_i)$? Our answer, which will become a principal tool of investigation, requires some preliminary notation.

Definition 3 i) P_n will denote the projection of $L^2([0,T]^n)$ onto $L_2(\widehat{[0,T]^n})$

$$(P_n h)(s_1 \cdot\cdot s_n) = \frac{1}{n!}\Sigma_{\pi\varepsilon S(n)} h(s_{\pi(1)}, \cdot\cdot s_{\pi(n)})$$

where S_n=permutation group on n letters.

ii) Let $0 \le k \le \min(m,n)$ $f\varepsilon L^2([0,T]^n)$, $g\varepsilon L^2([0,T]^m)$

$$(f\otimes_k(t)g)(s_1,\cdot\cdot,s_{n+m-2k}) \tag{6}$$
$$=\frac{1}{k!}\int_0^t\cdot\cdot\int_0^t f(r_1,\cdot\cdot,r_k,s_1,\cdot\cdot s_{n-k})g(r_1\cdot\cdot r_k,s_{n-k+1}\cdot\cdot s_{m+n-2k})dr_k\cdot\cdot dr_1$$

$$(f\theta_k(t)g)(s_1\cdot\cdot s_{m+n-2k}) = (P_{m+n-2k}[f\otimes(t)g])(s_1\cdot\cdot s_{m+n-2k}) \tag{7}$$

To illustrate, if $n \ge m = k$, then, using the symmetry of f and g, $(f\theta_m(t)g)(s_1\cdot\cdot s_{n-k}) \frac{1}{m!}\int_0^t\cdot\cdot\int_0^t f(r_1,\cdot\cdot r_m,s_1\cdot\cdot s_{n-m})$ $\times g(r_1\cdot\cdot r_m)dr_m\cdot\cdot dr_1$. It is useful to think of the functions f and g as tensors, for, in fact $L^2([0,T]^n)$ $\cong L^2[0,T]\otimes\cdot\cdot\otimes L^2([0,T])$ (n times). Therefore, as inspection of (6) and (7) suggests, $\theta_r(t)$ may be interpreted as a k-fold symmetrized tensor contraction.

Theorem 2. Let $f\varepsilon L^2([0,T]^n)$, $g\varepsilon L^2([0,T]^m)$. Then $I_t^{(n)}(f)I_t^{(m)}(g)\varepsilon L^2(F_t^b)$ for $t \le T$ and,

$$I_t^{(n)}(f)I_t^{(m)}(g) = \Sigma_{k=0}^{\min(m,n)} I_t^{(m+n-2k)}[(\tbinom{m+n}{m-k})^{2k} f\theta_k(t)g] \tag{8}$$

Before sketching a proof, let us first demonstrate that the l.h.s. of (8) is well-defined.

Lemma 1. Let $f\varepsilon L^2([0,T]^n)$, $g\varepsilon L^2([0,T]^m)$. For $t \le T$ $f\theta_k(t)g\varepsilon L^2([0,T]^{m+n-2k})$.

In fact $$\|f\theta_k(t)g\|^2_{m+n-2k} \le C\|f\|^2_n\|g\|^2_m. \tag{9}$$

where C depends on m, n and k.

Proof. Let $|S_n|$= cardinality (S_n) and $j=m+n-2k$.

Using Cauchy-Schwarz repeatedly:

$$\|f\theta_k(t)g\|_j \le \frac{|S(j)|}{(j!)^2}\Sigma_{\pi\varepsilon S(j)}\|f\otimes_r(t)g(s_{\pi(1)}\cdot\cdot s_{\pi(j)})\|_j^2$$

and

$$\|f\otimes_k(t)g(s_{\pi(1)}\cdot\cdot s_{\pi(j)})\|_j^2 = \frac{1}{(k!)^2}\int_0^T\cdot\cdot\int_0^T ds_{\pi(1)}\cdot\cdot ds_{\pi(j)}$$
$$\times[\int_0^t\cdot\cdot\int_0^t f(s_1\cdot\cdot s_r,s_{\pi(1)}\cdot\cdot)g(r_1\cdot\cdot r_r,\cdot\cdot s_{\pi(j)})dr_r\cdot\cdot dr_1]^2$$
$$\le \frac{1}{(k!)^2}\|f\|^2_m\|g\|^2_m$$

Thus $$\|f\theta_k(t)g\|_j \le [\frac{|S(j)|}{j!k!}]^2\|f\|^2_m\|g\|^2_m$$

Proof of theorem 2*: Only a sketch will be given, as details are involved and unrevealing. First, it suffices to treat the case when f and g are separable, since we can use lemma 1 to approximate general f and g by separable functions. This makes questions concerning the interchange of dt and db(t) integrations easy to resolve. The case n=1, m=1 follows directly by applying Ito's differentiation rule. Indeed, Ito's rule yields in general

$$I_t^{(n)}(f)I_t^{(m)}(g) = \int_0^t I_s^{(m)}(g)I^{(n-1)}(f(s,\cdot\cdot))db(s) + \tag{10}$$
$$\int_0^t I_s^{(n)}(f)I_s^{(m-1)}(g(s,\cdot\cdot))d_1b(s) + \int_0^t I_s^{(n-1)}(f(s,\cdot\cdot))$$
$$\times I_s^{(n-1)}(g(s\cdot\cdot))ds$$

Using (10) we can implement the following two stage induction scheme to prove the theorem for all m and n.

a) The case $(m,n)=(k,1)$ implies the case $(m,n)=(k+1,1)$
b) The cases $(m,n)=(k-1,j)$, $(k,j-1)$ and $(k-1,j-1)$ imply the case (k,j).

Equation (8), the multiplication formula, is actually a generalization of similar looking Hermite polynomial identity

$$h_m(x)h_n(x) = \Sigma_{k=0}^{\min(m,n)}((\tbinom{m}{k})(\tbinom{n}{k}))^{\frac{1}{2}}(\tbinom{r+q-2k}{r-k})^{\frac{1}{2}}h_{m+n-2k}(x) \tag{11}$$

where $h_n(x) = \frac{(-1)^n}{\sqrt{n!}}e^{-x^2/2}(d^n/dx^n)e^{x^2/2}$. To understand the connection, observe that the polynomials $h_m(x)$ provide an alternate means of constructing the decomposition of theorem 1. In fact, if $\{\phi_i\}_{i=1}^{\infty}$ is a complete orthonormal basis of $L^2([0,T])$ and $G_n \equiv \text{Span}\{\prod_{j=1}^{r} h_{p_j}(\int_0^t\phi_{i(j)}(s)db(s)) | p_1+\cdot\cdot+p_r=n, i_1,\cdot\cdot i_r$ pairwise unequal$\}$ then Ito [3] has shown that $H_n=\bar{G}_n$ ($^{-}$ denotes closure). (See also Kallianpur [4]). Thus a typical element $I_t^{(n)}(f)\varepsilon H_n$ is a generalization of a Hermite polynomial. The slight discrepency between the factors in (11) and (8) arises from the normalizations involved in the definitions of h_n, $I^{(n)}$ and $\otimes_k$.

(8) has consequences that relate directly to the theory of contractions on sums 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 kth order moments of the integral $I_t^{(m)}(f)$, and, indeed, a direct application of (8) using lemma 1 and a recursion argument yields:

Theorem 3 Let $n \ge 1$ and $f\varepsilon L^2(\widehat{[0,T]^n})$. For any $k \ge 1$, there exists $M_{m,k}>0$, independent of f, such that

$$E[I_T^{(n)}(f)]^{2k} \le M_{m,k}\|f\|_m^{2k}. \tag{12}$$

Now, Segal [5] has previously derived (12) by tensor product operator arguments, and, in addition, proves there exists a constant c such that M(m,k) may be replaced by k^{2ckn}. His argument thus connects (12) to a deeper general theory.

Theorem 3 has an interesting corollary.

Theorem 4 Let $\{f_m\}_{m=1}^{\infty}$ and f be functions in $L^2(\widehat{[0,T]^n})$. Then $\lim_{m\to\infty}\|f_m-f\|^2_n = 0$ iff $\lim_{m\to\infty} E[I_T^{(n)}(f_m)-I_T^{(n)}(f)]^2 = 0$ iff $\lim_{m\to\infty} E[I_T^{(n)}(f_m)-I_T^{(n)}(f)]^{2k} = 0$ for any or all k.

Proof By (4b) $E[I_T^{(n)}(f_m)-I_T^{(n)}(f)]^2 = \frac{1}{n!}\|f_m-f\|^2_n$. Using (12) completes the proof.

In the applications, we shall want to discuss multiple integrals not with respect to Brownian motion, but to an observation semi-martingale y_t. We again denote these integrals by $I_t^{(n)}(f)$ without explicitly indicating the dependence on y_t, which should be clear from the context of their use. The simplest definition of such an integral uses a result stated in theorem 5; namely, under condition (2) there exists a measure P_o on (Ω,F) such that i) y_t is Brownian on (Ω,F,P_o), and ii) P_o and P are mutually absolutely continuous.

Definition 4 For $f\varepsilon L^2([0,T]^n)$

$$I_t^{(n)}(f) = \int_0^t\cdot\cdot\int_0^{s(n-1)} f(s_1\cdot\cdot s_n)dy(s_n)\cdot\cdot dy(s_1)$$

is a r.v a.s. equal to the Brownian multiple integral defined on (Ω, F, P_o).

*Apparently, Japanese workers have also recently proved theorem 2 by means of functional analytic techniques due to Hida (Personal communication from T. Hida)

By absolute continuity, $I_t^{(n)}(f)$ is a well-defined r.v. on (Ω, F, P). Also, as further argument will show, $I_t^{(n)}(f)$ equals the iterated integral defined directly on (Ω, F, P) in the manner of definition 2, and thus the 'natural' interpretation of $I_t^{(n)}(f)$ as an iteration is preserved. It immediately follows from definition 4 that the multiplication formula holds for the observation semi-martingale case. Likewise, if $E_o(\frac{dP}{dP_o})^2<\infty$ then $E[I_T^{(n)}(f)]^{2k} = E_o\frac{dP}{dP_o}[I_T^{(n)}(F)]^{2k} \le (E_o(\frac{dP}{dP_o})^2)^{1/2}(E_o[I_T^{(n)}(f)]^{4k})^{1/2}$ shows that theorem 3 extends as well.

Finally, it is important to compute the mean and variance of integrals with respect to y_t.

Lemma 2 Let $E(\int_0^T h^2(x_s)ds)^n<\infty$. Then for $k\le m$ and $f\in L^2([0,T]^k)$,

i) $E[I_t^k(f)]^2 \le M_k \|f\|_k^2$ (M_r-does not depend on f)

ii) $E[I_t^k(f)] = \int_0^t..\int_0^{s(r-1)} f(s_1..s_r)E\{h(x_{s(1)})..h(x_{s(r)})\}ds(k)..ds(1)$

Proof The proof proceeds by induction on the order k, and the induction stops at k=m because of the condition $E(\int_0^t h^2(x_s)ds)^n<\infty$. Details will not be presented for lack of space.

Filter expansions and applications. We will now show that the Kallianpur-Striebel formula, (13), for the optimal estimate can be developed into a representation of the estimate as a ratio of two multiple integral expansions. This technique bears comparison to the work of Eterno [6], who derived simular expressions in an effort to approximate the conditional distribution of the signal given the observation process. Here we focus on the use of the expansion to derive equations for suboptimal filters.

Recall the filtering problem 1)-2). Denote $h(x(s))$ by $h(s)$,

$f_t(x(\cdot))$ by $f(t)$, and $E\{f(t)|F_t^y\}$ by $\widehat{f(t)}$, and define

$$L(t)=\exp[\int_0^t h(s)dy(s)-1/2\int_0^t h^2(s)ds].$$

Also, define a new measure P_o on (Ω,F) by

$$\frac{dP_o}{dP}= \exp[-\int_0^t h(s)dw(s)-1/2\int_0^t h^2(s)ds].$$

Theorem 5 Under the hypothese of (2)

i) P_o is a probability measure and P and P_o, are mutually absolutely continuous with $\frac{dP}{dP_o} = L(T)$.

ii) Under P_o $y(t)$ is a Brownian motion independent of $x(t)$.

iii) $x(t)$ has the same distribution under P_o as under P

iv) (Kallianpur, Striebel)

$$E\{f(t)|F_t^y\}=E_o\{f(t)\frac{dP}{dP_o}|F_t^y\}/E_o\{\frac{dP}{dP_o}|F_t^y\} = E_o\{f(t)L(t)|F_t^y\}/E_o\{L(t)|F_t^y\}. \quad (13)$$

Proof See Wong [7],

Let $L_{r-1}(t)=\int_0^t..\int_0^{s(r-1)} h(s_1)..h(s_r)L(s_r)dy(s_r)..dy(s_1)$.

Theorem 6 a) Partial expansion

Suppose $E[\int_0^t h^2(s)ds]^r<\infty$ and $E[f^2(s)(\int_0^t h^2(s)ds)^r]<\infty$,

Then $$\widehat{f(t)}=\frac{Ef(t)+\Sigma_{j=1}^r I_t^{(j)}[k_j]+E_o[f(t)L_T(t)|F_t^y]}{1+\Sigma_{j=1}^r I_t^{(j)}[1_j]+E_o[L_r(t)|F_t^y]}, \quad (14)$$

where $k_j(t, s_1,..s_j)=E[f(t)h(s_1)..h(s_j)]$ and

$1_j(s_1..s_j)=E[h(s_1)..h(s_j)]$

b) Full expansion. If $E[\exp[\int_0^t h^2(s)ds]]<\infty$ and $Ef^2(t)\exp\int_0^t h^2(s)ds<\infty$, then

$$f(t) = \frac{Ef(t)+\Sigma_{j=1}^\infty I_t^{(j)}[k_j]}{1+\Sigma_{j=1}^\infty I_t^{(j)}[1_j]}, \quad (15)$$

where k_j and 1_j are as above and the infinite series both converge in the $L^1(P)$ norm.

Remarks 1. The kernels k_j and 1_j depend only on the apriori 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 $\int_0^T h^2(s)ds$. Moreover

$$E_o(\frac{dP}{dP_o})^2 = E_oE_o\{\exp[2\int_0^T h(s)dy(s)-\int_0^T h^2(s)ds]|F_t^x\}$$
$$= E_o\exp[-\int_0^T h^2(s)ds]E_o\{\exp 2\int_0^T h(s)dy(s)|F_t^x\}$$
$$= E_o\exp[\int_0^T h^2(s)ds]$$

since $\int_0^T h(s)dy(s)$ conditioned on F_t^x is normal with variance $\int_0^T h^2(s)ds$.

3. Theorem 6 can be generalized without difficulty to vector valued processes.

Proof of theorem 6: For lack of space we give only an heuristic sketch. The principal idea comes from observing that, by using Ito's differentiation rule

$dL(t)=h(t)L(t)dy(t)$. In other symbols,

$$L(t)=1+\int_0^t h(s)L(s)dy(s) \quad (16)$$

By iterating (16):

$$L(t)=1+\int_0^t h(s)dy(s)+\int_0^t\int_0^s h(s)h(r)L(r)dy(r)dy(s).$$

Continuing such iteration ad infinitum we derive the formal expression

$$L(t)=1+\Sigma_{j=1}^\infty \int_0^t..\int_0^{s_{j-1}} h(s_1)..h(s_j)dy(s_j)..dy(s_1). \quad (17)$$

Now substitute (17) into the term $E_o[L(t)|F_t^y]$. We get:

$$E_o[L(t)|F_t^y]=1+\Sigma_{j=1}^\infty E_o\{\int_0^t..\int_0^{s_{j-1}} h(s_1)..h(s_j)dy(s_j)..dy(s_1)|F_t^y\}$$
$$=1+\Sigma_{j=1}^\infty\int_0^t..\int_0^{s_{j-1}} E_o\{h(s_1)..h(s_j)|F_t^y\}dy(s_j)..dy(s_1)$$
$$=1+\Sigma_{j=1}^\infty\int_0^t..\int_0^{j-1} E_o\{h(s_1)..h(s_j)\}dy(s_j)..dy(s_1)$$
$$=1+\Sigma_{j=1}^\infty I_t^{(j)}(1_j). \quad (18)$$

The second equality uses a stochastic 'Fubini' theorem found, for example, in Liptser and Shiryayev [8]; for the process $\phi(s)$ adapted to the Brownian motion (b_t, F_t) and satisfying $E[\int_0^T\phi^2(s)ds]<\infty$

$$E\{\int_0^T\phi_s db(s)|F_t^b\} = \int_0^T E[\phi_s|F_s^b]db(s).$$

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

$$E_o\{f(t)L(t)|F_t^y\}=1+\Sigma_{j=1}^\infty I_t^{(j)}(k_j). \quad (19)$$

Now (18) and (19) can be substituted in

$$f(t) = E_o\{f(t)L(t)|F_t^y\}/E_o\{L(t)|F_t^y\}$$

to formally derive theorem 6, b). The partial expansion if proved by carrying out the iteration procedure of (16) only a finite number of times. The various hypotheses in theorem 6 merely guarantee that the steps in each calculation are valid.

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 $y(\cdot)$ 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

$$c(t)=c_o(t)+\Sigma_{n=1}^{r} I_t^{(n)}(d_n(t))$$

with $c_n(t)\epsilon L^2(\widehat{[0,T]}^n)$ is called an rth order expansion of $y(\cdot)$.

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

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

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

The best rth estimate will be denoted by $f(t)$ (with r understood), and its kernels by $a_o, a_1, .. a_r$.

Given an order r, how can we find $f(t)$, that is how can we determine $a_o, a_1, .., a_r$? 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 $f(t)E_o[L(t)|F_t^y]$ of the estimate with the denominator of (13). If

$$E[\int_0^T h^2(s)ds]^{2r}<\infty$$

then the expansion of $E_o[L(t)|F_t^y]$ in (14) applies, and

$$f(t)E_o[L(t)|F_t^y]=[a_o(t)+\Sigma_{n=1}^{r} I_t^{(n)}(a_n(t))][1+\Sigma_{n=1}^{2r} I_t^{(n)}(1_n)$$
$$+E_o\{L_{2r}(t)|F_t^y\}]=g_o(t)+\Sigma_{n=1}^{3r} I_t^{(n)}(g_n(t))+f(t)E_o\{L_{2r}(t)|F_t^y\}. \quad (20)$$

The g_n, n=1,..3r are calculated from $a_n(t)$ and $1_n(t)$ by use of the multiplication formula.

Theorem 7. Suppose $E[\int_0^T h^2(s)ds]^{2r}<\infty$, $Ef^2(t)<\infty$ and $Ef^2(t)[\int_0^T h^2(s)ds]^{2r}<\infty$.

Then $\tilde{f}(t)=a_o(t)+\Sigma_{n=1}^{r} I_t^{(n)}(a_n(t))$ is the best rth order estimate iff

$$g_o(t)=Ef(t) \quad (21)$$

$$g_n(t,s_1..s_n)=E\{f(t)h(s_1)..h(s_n)\}=k_n,\ 1\leq n\leq r.$$

Proof: We must show that

$$E[f(t)-f(t)]^2\leq E[f(t)-c(t)]^2 \quad (22)$$

for all nth order expansions $c(t)$ iff (21) holds. Recall that $\hat{f}(t)$ may be interpreted as the projection of $f(t)$ onto $L^2(\Omega, F_t^y, P)$. Thus the projection theorem implies

$$E[f(f)-\tilde{f}(t)]^2=E[f(t)-\hat{f}(t)]^2+E[\hat{f}(t)-\tilde{f}(t)]^2$$
$$+2E[f(t)-\hat{f}(t)][\hat{f}(t)-\tilde{f}(t)]=E[f(t)-\hat{f}(t)]^2+E[f(t)-\tilde{f}(t)]^2$$

Applying this calculation to the r.h.s. of (22) also, (22) holds iff

$$E[\hat{f}(t)-\tilde{f}(t)]^2\leq E[\hat{f}(t)-c(t)]^2 \quad (23)$$

for all $c(t)$. But according to lemma 2, the set of rth order expansions in $y(\cdot)$ is a Hilbert space, and thus, applying the projection theorem again, (23) holds iff

$$E[\hat{f}(t)-\tilde{f}(t)]c(t)=0 \quad (24)$$

for all rth order expansions $c(t)$. Now substitute the expression (13) for $f(t)$ into (24):

$$E[\hat{f}(t)-\tilde{f}(t)]c(t)=E[\frac{E_o\{f(t)L(t)F_t^y\}-\tilde{f}(t)E_o\{L(t)|F_t^y\}}{E_o\{L(t)|F_t^y\}}c(t)$$
$$=E\,E[\frac{dP_o}{dP}\ F_t^y][E_o\{f(t)L(t)|F_t^y\}-\tilde{f}(t)E_o\{L(t)|F_t^y\}]c(t)$$
$$=E_o[E_o\{f(t)L(t)|F_t^y\}-\tilde{f}(t)E_o\{L(t)|F_t^y\}]c(t)] \quad (25)$$

The second equality in (25) uses the identities

$$E\{\frac{dP_o}{dP}|F_t^y\}=[E_o\{\frac{dP}{dP_o}|F_t^y\}]^{-1}$$
$$=[E_o\{L(t)|F_t^y\}]^{-1},$$

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

$$E_o\{\hat{f}(t)L(t)|F_t^y\}=Ef(t)+\Sigma_{n=1}^{r} I_t^{(n)}(k_n)+E_o\{f(t)L_r(t)|F_t^y\}$$

in (25),

$$E[\tilde{f}(t)-f(t)]c(t)=E_o[Ef(t)-g_o+\Sigma_{n=1}^{r} I_t^{(n)}(k_n-g_n)]c(t)$$
$$+E_o[f(t)c(t)E_o\{L_{2r}(t)|F_t^y\}] \quad (26)$$
$$+E_o[c(t)E_o\{f(t)L_r(t)|F_t^y\}]$$

An application of lemma 2 show 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 $c(t)$ iff $k_n=g_n$ 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 rth order estimate, since the $g_n(t)$, $0\leq n\leq r$, are found from $a_n(t)$, $0\leq n\leq r$, by the formula (8). To illustrate, if r=1, $1_1(s)=Eh(s)$ and

$$Ef(t)=g_o(t)=a_o(t)+\int_0^t a_1(t,u)Eh(u)dw$$
$$Ef(t)h(s)=g_1(t,s)$$
$$=a_o(t)Eh(s)+a_1(t,s).$$

Solving for $a_o(t)$,

$$a_1(t,s)+\int_0^t a_1(t,u)\mathrm{cov}[h(s),h(u)]du=\mathrm{cov}[f(t),h(s)].$$

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 $Eh(s)\equiv 0$, $Ef(t)\equiv 0$ for simplcity,

$$a_o(t)=-\int_0^t\int_0^{u_1} a_2(t,u_1,u_2)Eh(u_1)h(u_2)ds_2ds_1 \quad (27a)$$

$$a_1(t,s)=Ef(t)h(s)-\int_0^t a_1(t,u)Eh(s)h(u)du$$
$$-\int_0^t\int_o^{u_1} a_2(t,u_1,u_2)\mathrm{cov}[h(s),h(u_1)h(u_2)]du_2du_1 \quad (27b)$$

$$a_2(t,s_1,s_2)=\mathrm{cov}[f(t),h(s_1),h(s_2)]$$
$$-\int_0^t a_1(t,u)\mathrm{cov}[h(s_1),h(s_2),h(u)]du$$
$$-\int_0^t[a_2(t,s_1,u)Eh(s_2h(u)+a_2(t,s_2,u)Eh(s_1)h(u)]du$$
$$-\int_0^t\int_o^{u_1} a_2(t,u_1,u_2)\mathrm{cov}[h(s_1),h(s_2),h(u_1)h(u_2)]du_2du_1 \quad (27c)$$

[In (27), cov $[X_1,..X_r] \equiv E(X_1-EX_1)\ldots(X_r-EX_r)$.]
(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 a_2. For fixed t this equation is of Fredholm type for $a_2(t,\cdot,\cdot)$ 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

$$\hat{f}(t)=\int_0^t a(t,s)dy(s)$$

where

$$a_1(t,s)+\int_0^t a_1(t,u)Eb(u)b(s)=Eb(t)b(s). \tag{28}$$

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

$$\hat{f}(t)=\int_0^t a(t,s)dy(s)=Ef(t)+\Sigma_{j=1}^{\infty} I_t^{(j)}(k_j)/1+\Sigma_{j=1}^{\infty} I_t^{(j)}(1_j) \tag{29}$$

or

$$\int_0^t a(t,s)dy(s)[1+\Sigma_{j=1}^{\infty} I_t^{(j)}(1_j)]=Ef(t)+\Sigma_{j=1}^{\infty} I_t^{(j)}(k_j). \tag{30}$$

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

$$ja(t,\cdot)\theta_o(t)1_{j-1}+a_1(t,\cdot)\theta_1(t)1_{j+1}=k_j. \tag{31}$$

It can now be shown that if (31) is satisfied for j=1, it is satified for all $j\geq 1$, a result following from the identity for Gaussian random variables:

$$1_j(s_1..s_j)=Eb(s_1)..b(s_j)$$

$$=\Sigma_{j=2}^{n}\text{cov}(b(s_1)b(s_2))E[b(s_2)..b(s_{j-1})b(s_{j+1})..b(s_j)]$$

(see Miller [9]. This derivation is somewhat formal because the condition for the full expansion in (29) to hold is that $E[\exp\int_0^t b^2(s)ds]<\infty$, which is valid only for small t.

4. Relationship to Second Quantization (After Segal and Nelson).

Let H be a real Hilbert space and let $F:H\to RV(\Omega,A,u)$ be the unit Gaussian determined random field. If $f_1..,f_n$ are orthonormal in H and ϕ is a Bounded Baire function on R^n, then

$$\int_\Omega \phi(F(f_1),..,F(f_n))=\frac{1}{(2\pi)^{n/2}}\int_{R^n}\phi(x)e^{-\|x\|^2/2}dx$$

For concreteness (Ω,A,μ) may be chosen to be countably infinite copies of $(R,B(R),(2\pi)^{-1/2}e^{-x^2/2}dx)$.

If E donotes expectation on (Ω,A,μ) then

$$E(F(f_1)..F(f_{2n+1}))=0 \tag{32}$$

$$E(F(f_1)..F(f_{2n}))=\Sigma[f_{i_1},f_{j_1}]\ldots[f_{i_n},f_{j_n}] \tag{33}$$

Where the sum is over all pairings of $1,\ldots,2_n$, i.e. $i_1<\ldots<i_n$; $i_1<j_1,\ldots,i_n,j_n$, and $(i_1,j_1,\ldots,i_n,j_n)$ is a permutation $1,\ldots 2_n$.

$L^p(\Omega,A,\mu)$ is denoted by $L^p(H)$ and $\Gamma(H)$ denotes $L^2(H)$. $\Gamma(H)\leq n$ be the closed linear span in $\Gamma(H)$ of all elements of the form $F(f_1)\ldots F(f_n)$ with $m\leq n$ and let $\Gamma(H)_n$ denote the orthogonal complement of $\Gamma(H)_{\leq n-1}$ in $\Gamma(H)_{\leq n}$. For $f_1,..,f_n$ in H we define the Wick polynomial:

$$:F(f_1)\ldots F(f_n):$$

to be the orthogonal projection of $F(f_1)\ldots F(f_n)$ into $\Gamma(H)_n$. In the special case, where H is one dimensional and hence $\Gamma(H)=L^2(R,B(R),(2\pi)^{-1/2}e^{-x^2/2}dx)$, $\Gamma(H)_n$ is the one dimensional subspace spanned by the nth Hermite polynomial and $:x^n:$ is the nth. Hermite polynomial normalization so that the leading coefficient is 1.

We have the formula

$$[:F(f_1)..F(f_n):,:F(g_1)..F(g_n):]$$

$$=\Sigma_\pi[f_{\pi(1)},g_1]..[f_{\pi(n)},g_n]. \tag{34}$$

where the sum is over all permutations π of 1,...n. If all the f's and g's are equal, we get

$$[:F(f)^n:,F(f)^n]=\frac{1}{2^{1/2}}\int_{-\infty}^{\infty}(:x^n:)^2e^{-x^2/2}dx=n!. \tag{35}$$

Let H_1 be the complexification of H (and let H_n denote the n-fold Hilbert space) symmetric tensor product of H_1 with itself. On H_2 we define the inner product such that

$$[\text{Sym}(f_1\otimes\ldots\otimes f_n),\text{Sym}(g_1\otimes\ldots g_n)]=\Sigma_\pi[f_{\pi(1)},g_1]\ldots[f_{\pi(n)},g_n] \tag{36}$$

where

$$\text{Sym}\ (f_1\ldots\otimes f_n)=\frac{1}{n!}\Sigma_\pi f_{\pi(1)}\otimes\ldots\otimes f_{\pi(n)}. \tag{37}$$

From (34) and (36), that the mapping

$$:F(f_1)\ldots F(f_n):\ \text{Sym}(f_1\otimes\ldots\otimes f_n)$$

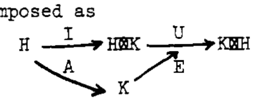
extends uniquely to the unitary operator from $\Gamma(H)_n$ onto H_n. We use this mapping to identify $\Gamma(H)_n$ and H_n. Analogous to the fact that the one-dimensional Hermite polynomials span $L^2(R,B(R),(2\pi)^{-1/2}e^{-x^2/2}dx)$, Segal proved

$$\Gamma(H)=\Sigma_{n=o}^{\infty}H_n, \tag{38}$$

for arbitrary real Hilbert space H. $\Gamma(H)$ is Fock Space.

If the random field $F(f)=\int f dB$, where $f\varepsilon L^2(R)=H$ and B is the standard Wiener process, the elements of $\Gamma(H)_n$ are multiple Wiener integrals (in the sense of Ito).

The space $\Gamma(H)$ is intrinsically attached to the structure of H as a real Hilbert space. Thus if $U:H\to K$ is an orthogonal mapping of one real Hilbert space into another, it induces a unitery mapping $\Gamma(U):\Gamma(H)\to\Gamma(K)$, where on H_n, $\Gamma(U)=\underbrace{U\otimes\ldots\otimes U}_{n\text{-fold}}$. Similarly if $I:H\to K$ is an isometric embedding then it induces an isometric embedding $\Gamma(I)$: $\Gamma(H)\to\Gamma(K)$ and similarly for an orthogonal projection $E:H\to K$. If $A:H\to K$ is a contraction then $\Gamma(A):\Gamma(H)\to\Gamma(K)$ is defined to be the direct sum to $\Gamma(A)_n:H_n\to K_n$, where $\Gamma(A)_n=\underbrace{A\otimes\ldots\otimes A}_{n\text{-fold}}$. Now any contraction $A:H\to K$ can be decomposed as



where I, U and E are as above.

Hence $\Gamma(A)=\Gamma(E)\Gamma(U)\Gamma(I)$. Now $\Gamma(A)$ is doubly Markovian in the sense that

$$\alpha\geq 0\to\Gamma(A)\alpha\geq 0$$
$$\Gamma(A)1=1$$
$$E\Gamma(A)\alpha=E\alpha. \tag{39}$$

Any doubly Markovian operation is a contraction from L^p to L^p.

It turns out that $\Gamma(A)$ has stronger contractive 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 $L^2(\Omega,A,u)$ can be characterised as linear positivity preserving operators which are idempotant, of norm ≤ 1 and preserve constants. We also know that for p $[1,\infty]$, $p\neq 2$, all linear operators T on $L^p(\Omega,A,u)$, which are idempotent, contracting and such that T1=1 is necessarily a conditional expectation.

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

$$\|A\|\leq\left(\frac{q-1}{p-1}\right)^{1/2} \tag{40}$$

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

[1] M. Fujisaki, G. Kallianpur and H. Kunita, "Stochastic Differential Equations for the Nonlinear Filtering Problem," Osaka J. Math., Vol. 9, 1972, pp. 19-40.

[2] R.W. Brockett, "Volterra Series and Geometric Control Theory," Automatica, 12 (1976), pp. 167-176.

[3] K. Ito, "Multiple Wiener Integrals," J. Math. Soc. Japan, 13 (1951) pp. 157-169.

[4] G. Kallianpur, "The Role of Reproducing Kernel Hilbert Spaces in the Study of Gaussian Processes," Advances in Probability, Vol. 2, ed. P. Ney, Marcel Dekker, N.Y., 1970.

[5] Segal, "Construction of Non-Linear Local Quantum Processes: I," Ann. Math., 92 L970), pp. 462-481.

[6] J.S. Eterno, Nonlinear Estimation Theory and Phase-Lock Loops, Ph.D. thesis, MIT, 1976.

[7] E. Wong, Stochastic Processes in Information and Dynamical Systems, McGraw Hill, New York, 1971.

[8] R.S. Liptser and A.N. Shiryayen, Statistics of Random Processes, Vol. 1, Springer-Verlag, New York, 1979.

[9] K.S. Miller, Multidimensional Gaussian Distributions, John Wiley, New York 1965.

Acknowledgment

This research has been supported by the Air Force Office of Scientific Research under Grant No. AFOSR-77-3281B