MULTIPLE INTEGRAL EXPANSIONS FOR NONLINEAR FILTERING

S. K. Mitra 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 developement 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:

\[ y_t = \int_0^T h(x_s) ds + \epsilon_t \quad (1) \]

under the assumption that \( x_t \) and \( \epsilon_t \) are independent processes with \( \epsilon_t \) being a standard Brownian motion.

If \( f_t(x_t) = \int_0^T h_t(x_s) ds \) is a causal functional of the signal \( x_t \) and \( h_t \) is a sub-\( \sigma \)-algebra generated by \( x_t \) up to \( s \), then we are interested in calculating the optimal least squares estimate of \( f_t(x_t) \)

\[ E[f_t(x_t)|F_t] \quad \text{for} \quad t \geq 0. \]

Definition 1: \( y_t \) defined in (1) is called an observation semi-martingale.

Throughout, let \((H_t,F_t)\) denote the underlying probability space.

Now \( E[f_t(x_t)|F_t] \) is a measurable function of \( x_t \).

3. 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:

\[ y_t = \int_0^T h(x_s) ds + \epsilon_t \quad (1) \]

under the assumption that \( x_t \) and \( \epsilon_t \) are independent processes with \( \epsilon_t \) being a standard Brownian motion.

If \( f_t(x_t) = \int_0^T h_t(x_s) ds \) is a causal functional of the signal \( x_t \) and \( h_t \) is a sub-\( \sigma \)-algebra generated by \( x_t \) up to \( s \), then we are interested in calculating the optimal least squares estimate of \( f_t(x_t) \)

\[ E[f_t(x_t)|F_t] \quad \text{for} \quad t \geq 0. \]

Definition 1: \( y_t \) defined in (1) is called an observation semi-martingale.

Throughout, let \((H_t,F_t)\) denote the underlying probability space.

Now \( E[f_t(x_t)|F_t] \) is a measurable function of \( x_t \).

4. 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:

\[ y_t = \int_0^T h(x_s) ds + \epsilon_t \quad (1) \]

under the assumption that \( x_t \) and \( \epsilon_t \) are independent processes with \( \epsilon_t \) being a standard Brownian motion.

If \( f_t(x_t) = \int_0^T h_t(x_s) ds \) is a causal functional of the signal \( x_t \) and \( h_t \) is a sub-\( \sigma \)-algebra generated by \( x_t \) up to \( s \), then we are interested in calculating the optimal least squares estimate of \( f_t(x_t) \)

\[ E[f_t(x_t)|F_t] \quad \text{for} \quad t \geq 0. \]

Definition 1: \( y_t \) defined in (1) is called an observation semi-martingale.

Throughout, let \((H_t,F_t)\) denote the underlying probability space.

Now \( E[f_t(x_t)|F_t] \) is a measurable function of \( x_t \).

5. 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:

\[ y_t = \int_0^T h(x_s) ds + \epsilon_t \quad (1) \]

under the assumption that \( x_t \) and \( \epsilon_t \) are independent processes with \( \epsilon_t \) being a standard Brownian motion.

If \( f_t(x_t) = \int_0^T h_t(x_s) ds \) is a causal functional of the signal \( x_t \) and \( h_t \) is a sub-\( \sigma \)-algebra generated by \( x_t \) up to \( s \), then we are interested in calculating the optimal least squares estimate of \( f_t(x_t) \)

\[ E[f_t(x_t)|F_t] \quad \text{for} \quad t \geq 0. \]

Definition 1: \( y_t \) defined in (1) is called an observation semi-martingale.

Throughout, let \((H_t,F_t)\) denote the underlying probability space.

Now \( E[f_t(x_t)|F_t] \) is a measurable function of \( x_t \).


The case \((m,n)=(k,l)\) implies the case \((m,n)=(k+1,l)\).

b) The cases \((m,n)=(k-1,l), (k,j-1)\) 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

\[
h_n(x) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k e^{-x^2/2} \frac{d^k}{dx^k} e^{x^2/2}.
\]

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

\[
\sum_{i=1}^{\infty} \phi_i(s) dB_i(t) = \int_0^T \phi_t(s) dB_s(t),
\]

then Itô [3] has shown that \(h_n(x) = \int_0^T e^{-x^2/2} \frac{d^n}{dx^n} e^{x^2/2} \). The slight discrepancy between the factors in (11) and (8) arises from the normalizations involved in the definitions of \(h_n\) and \(\phi_i\).

(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 \(k\)th order moments of the integral \(I_t(f)\), and, indeed, a direct application of (8) using lemma 1 and a recursion argument yields:

\[
I_t^n(f) = 0 \iff \lim_{k \to \infty} \frac{1}{k} I_t^n(f) = 0.
\]

Theorem 3 has an interesting corollary. Theorem 4 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 \(k^2 c m^n\). His argument thus connects (12) to a deeper general theory.

Theorem 5 has an interesting corollary. 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 3; namely, under condition (2) there exists a measure \(P_0\) on \(\mathbb{R}, \mathcal{F}_0\) such that \(1)\) \(Y_t\) is Brownian on \((\mathbb{R}, \mathcal{F}_0, P_0)\), and \(2)\) \(P_0\) and \(P_0\) are mutually absolutely continuous.

Definition 1 For \(f \in L^2([0,T])\)

\[
I_t^n(f) = \int_0^T \frac{d^n}{ds^n} f(s, \ldots, s_n) dB_s(t),
\]

is a r.v.s. equal to the Brownian multiple integral (8).

\[
I_t^n(f) = \frac{1}{n!} \int_0^T \frac{d^n}{ds^n} f(s, \ldots, s_n) dB_s(t).
\]

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

\[
\int_0^T \frac{d^n}{ds^n} f(s, \ldots, s_n) dB_s(t) = \int_0^T \frac{d^n}{ds^n} f(s, \ldots, s_n) dB_s(t).
\]
By absolute continuity, \( I^n_t(f) \) is a well-defined r.v. on \((\Omega, \mathcal{F}, P)\). Also, as further argument will show, \( I^n_t(f) \) equals the iterated integral defined directly on \((\Omega, \mathcal{F}, P)\) in the manner of definition 2, and thus the 'natural' interpretation of \( I^n_t(f) \) as an iteration is preserved. It immediately follows from definition 4 that the multiplication formula holds for the observation semi-martingale case. Likewise, if \( F_0 \) then \( \mathbb{E}_0 \{ \frac{dP_0}{dP} \}^2 < \infty \) shows that theorem 3 extends as well.

Finally, it is important to compute the mean and variance of integrals with respect to \( \gamma_t \).

**Remark 1.** The kernels \( k \) and \( l \) depend only on the a priori distribution of \( x(-\cdot) \).

**Theorem 5**

Under the hypotheses of (2)

1) \( P_0 \) is a probability measure and \( P \) and \( P_0 \) are mutually absolutely continuous with \( \frac{dP_0}{dP} = L(t) \).

2) Under \( P_0 \) \( y(t) \) is a Brownian motion independent of \( x(t) \).

3) \( x(t) \) has the same distribution under \( P_0 \) as under \( P \) iv) (Kallianpur, Striebel, 1975)

\[
\mathbb{E}_0 \{ f(t) | \gamma_t \} = \int f(t) \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \tag{13}
\]

**Proof**

See Wong (1977).

Let \( L_{t-s} = \mathbb{E}_0 \{ f(t) | \gamma_t \} \),

\[
\text{Theorem 6 a) Partial expansion}
\]

Suppose \( \mathbb{E}_0 \{ h^2(s) ds \} < \infty \) and \( \mathbb{E}_0 \{ h^2(s) ds | \gamma_s \} < \infty \).

Then \( \mathbb{E}_0 \{ f(t) | \gamma_t \} - \mathbb{E}_0 \{ f(t) | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \tag{14}
\]

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 \( k \) and \( l \) depend only on the a priori distribution of \( x(t) \).

2. The condition \( \mathbb{E}_0 \{ \exp \{ f(t) \} | \gamma_t \} \leq \infty \) places strong restrictions on the growth of the moments of \( h^2(s) ds \).

Finally, it is important to compute the mean and variance of integrals with respect to \( \gamma_t \).

**Lemma 2**

Let \( \mathbb{E}_0 \{ \gamma_t^2(x_s^2) ds < \infty \) Then for \( \mathbb{E}_0 \{ f(t) | \gamma_t \} \),

\[
\mathbb{E}_0 \{ f(t) \mathbb{E}_0 \{ f(t | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

**Proof**

The proof proceeds by induction on the order \( k \), and the induction stops at \( k = m \) because of the condition \( \mathbb{E}_0 \{ f(t) \} < \infty \). Details will not be presented for lack of space.

**Filter expansions and applications.** We will now show that the Kallianpur-Striebel formula, (11), 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 [1], who derived similar 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 \( h(s) \),

\[
\mathbb{E}_0 \{ f(t) | \gamma_t \} = f(t) \mathbb{E}_0 \{ f(t | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

Also, define a new measure \( P \) on \((\Omega, \mathcal{F})\) by \( \frac{dP_0}{dP} = L(t) \),

\[
\text{Theorem 6 b) Full expansion. If } \mathbb{E}_0 \{ f(t) \mathbb{E}_0 \{ f(t | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \},
\]

\[
\mathbb{E}_0 \{ f(t) \mathbb{E}_0 \{ f(t | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

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 \( k \) and \( l \) depend only on the a priori distribution of \( x(t) \).

2. The condition \( \mathbb{E}_0 \{ \exp \{ f(t) \} | \gamma_t \} \leq \infty \) in (6) places strong restrictions on the growth of the moments of \( h^2(s) ds \). Moreover

\[
\mathbb{E}_0 \{ f(t) \mathbb{E}_0 \{ f(t | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

since \( \mathbb{E}_0 \{ h(s) \gamma_t \} \) conditioned on \( \gamma_t \) is normal with variance \( \mathbb{E}_0 \{ h^2(s) ds \} \).

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

\[
L(t) = h(t) L(t) dy(t)
\]

In other symbols,

\[
L(t) = h(t) L(t) dy(t)
\]

By iterating (16):

\[
L(t) = h(t) L(t) dy(t)
\]

Continuing such iteration ad infinitum we derive the formal expression

\[
L(t) = \mathbb{E}_0 \{ f(t) | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

Now substitute (17) into the term \( \mathbb{E}_0 \{ f(t) | \gamma_t \} \). We get:

\[
\mathbb{E}_0 \{ f(t) \gamma_t \} = \mathbb{E}_0 \{ f(t) \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

The second equality uses a stochastic 'Fubini' theorem found, for example, in Liptser and Shiryayev [9]; for the process \( f(s) \) adapted to the Brownian motion \( (b_t, \mathcal{F}_t) \) and satisfying \( \mathbb{E}_0 \{ f(s) ds \} \), and the fourth equality by definition. By a similar calculation,

\[
\mathbb{E}_0 \{ f(t) | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

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

\[
\mathbb{E}_0 \{ f(t) \gamma_t \} = \mathbb{E}_0 \{ f(t) | \gamma_t \} | \gamma_t | \frac{dP_0}{dP} | \gamma_t | \frac{dP_0}{dP} \}
\]

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 simple 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, it is possible to construct the truncated filter. However, this design 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

\[ c(t) = a_0(t) + \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \]

with \( a_n(t) \in L^2(0,T) \) is called an \( r \)-th order expansion of \( y(\cdot) \).

b) An \( r \)-th order expression \( a(t) \), satisfying

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

for any other \( r \)-th order \( c(t) \), is called the best \( r \)-th order estimate of \( f(t) \). The best \( r \)-th estimate will be denoted by \( f(t) \) (with \( r \) understood), and its kernels by \( a_0, a_1, \ldots, a_r \).

Given an order \( r \), how can we find \( f(t) \)? As it turns out, we can apply the multiplication formula to the filter expansion to write integral equations for the kernels \( a_n(t) \). Begin by considering the product of the estimate with the denominator of (13). If

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

then the expansion of \( E_o[ L(t) | Y_t] \) in (14) applies, and

\[ f(t) = a_0(t) + \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \]

and

\[ E_o[ L(t) | Y_t] a_0(t) = \frac{E_o[f(t)]}{E_o[\mathcal{L}(t) | Y_t]} \]

The \( a_n(t) \), \( n=1, \ldots, r \), are calculated from \( a_0(t) \) and \( 1(t) \) by use of the multiplication formula.

Theorem 7. Suppose \( E_o[ L^2(s) ds]^2 < \infty \), \( E_o[ L^2(s) ds] \) and \( E_o[ f^2(t) ds] \) are finite, and \( a_0(t) = \frac{E_o[f(t)]}{E_o[\mathcal{L}(t) | Y_t]} \). Then

\[ f(t) = a_0(t) + \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \]

is the best \( r \)-th order estimate iff

\[ a_0(t) = E(f(t)) \]

(21)

\[ a_n(t,s_1, s_2) = E(f(t)h(s_1)h(s_2)) = k_n(s_1,s_2), \quad 1 \leq n \leq r. \]

Proof: We must show that

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

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

\[ E[f(t) - a(t)]^2 = E[f(t) - f(t)]^2 + E[f(t) - f(t)]^2 \]

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

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

(23)

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

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

(24)

for all \( r \)-th order expansions \( c(t) \). Now substitute the expression (13) for \( f(t) \) into (24):

\[ E[f(t) - f(t)]^2 = E \left[ \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \right]^2 \]

\[ = E \left[ a_0(t) L(t) \right]^2 + E \left[ \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \right]^2 \]

\[ = E \left[ a_0(t) L(t) \right]^2 + E \left[ \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \right]^2 \]

The second equality in (25) uses the identities

\[ E[f(t) - f(t)]^2 = E \left[ \sum_{n=1}^{r} a_n(t) (\delta_n(t)) \right]^2 \]

which are easily demonstrated. Now under \( F_0, y(\cdot) \) is a Brownian motion and integrals of different orders are orthogonal. Thus, using (20) and

\[ E[f(t) - f(t)]^2 = \sum_{n=1}^{r} \left( k_n(s_1,s_2) \right)^2 \]

in (25),

\[ E[f(t) - f(t)]^2 = \sum_{n=1}^{r} \left( k_n(s_1,s_2) \right)^2 \]

(26)

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 \( k_n(s_1,s_2) \) and this completes the proof.

The equations (26) are actually integral equations for the kernels \( a_n(t) \) of the best \( r \)-th order estimate, since the \( k_n(s_1,s_2) \), are found from \( a_n(t) \) by the formula (8). To illustrate, if \( r=1 \), \( a_1(t,s) = E[h(s)|L(t)] \) and

\[ E[f(t) - f(t)]^2 = \sum_{n=1}^{r} \left( k_n(s_1,s_2) \right)^2 \]

(27)

Solving for \( a_0(t) \),

\[ a_0(t) = E[f(t)] \]

(21)

\[ a_1(t,s) = E[h(s)|L(t)] \]

(22)

This is the familiar Wiener-Hopf equation for the best linear estimate. In the best quadratic case, the equations become more complicated. They are, assuming \( E[h(s)|L(t)] \) for simplicity,

\[ a_0(t) = \int_0^T a_1(t,u) E[h(u)|L(t)] du \]

(27a)

\[ a_1(t,s) = \int_0^T a_1(t,u) E[h(u)|L(t)] du \]

(27b)

The equations (27) are actually integral equations for the kernels \( a_n(t) \) of the best \( r \)-th order estimate, since the \( k_n(s_1,s_2) \), are found from \( a_n(t) \) by the formula (8). To illustrate, if \( r=1 \), \( a_1(t,s) = E[h(s)|L(t)] \) and

\[ E[f(t) - f(t)]^2 = \sum_{n=1}^{r} \left( k_n(s_1,s_2) \right)^2 \]

(26)

Now solving for \( a_0(t) \),

\[ a_0(t) = \int_0^T a_1(t,u) E[h(u)|L(t)] du \]

(27a)

\[ a_1(t,s) = \int_0^T a_1(t,u) E[h(u)|L(t)] du \]

(27b)

This is the familiar Wiener-Hopf equation for the best linear estimate. In the best quadratic case, the equations become more complicated. They are, assuming \( E[h(s)|L(t)] \) for simplicity,
In (27), \[ \text{cov} \{X_1,\ldots,X_n\} = \text{E}(X_1 - \text{E}X_1)(X_n - \text{E}X_n). \]

(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 \( a_0 \). For fixed \( t \) this equation is of Fredholm type for \( a_2(t,\cdot) \) 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

\[
\begin{align*}
 dx(t) &= db(t) \\
 dy(t) &= x(t) dt + dw(t)
\end{align*}
\]

where \( b(\cdot) \) and \( w(\cdot) \) are independent Brownian motions. Then we can show that the optimal filter is

\[
\hat{a}(t)+\hat{a}(t,u)dy(s)
\]

where \( a_0(t,s) \) is the optimal filter. We shall not go into this theory here.

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

\[
j a(t) = a(t,s) dy(s) + n a(t,u) dy(s).
\]

By expanding the r.h.s of (29) using (8), we get

\[
\begin{align*}
 j(t) &= j(t,s) dy(s) + n j(t,u) dy(s) \\
 a(t) &= a(t,s) dy(s) + n a(t,u) dy(s)
\end{align*}
\]

where \( a_0(t,s) \) is the optimal filter. We shall not go into this theory here.

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

Let \( H \) be a real Hilbert space and let \( F:H \to \mathbb{E}(\Omega,A,u) \) be the unit Gaussian determined random field. If \( f_1,\ldots,f_n \) are orthonormal in \( H \) and \( \hat{a} \) is a Bounded Baire function on \( H \), then

\[
\text{E}(F(f_1)\ldots F(f_n)) = \int_{\Omega} F(f_1)\ldots F(f_n) d\mu = \text{E}(F(f_1)\ldots F(f_n))
\]

For concreteness, \( (\Omega,A,u) \) may be chosen to be countably infinite copies of \( \Omega \), \( \mathbb{E}(\Omega,A,u) \) is the natural Gaussian determined random field. If \( f_1,\ldots,f_n \) are orthonormal in \( H \) and \( \hat{a} \) is a Bounded Baire function on \( H \), then

\[
\text{E}(F(f_1)\ldots F(f_{2n})) = \int_{\Omega} F(f_1)\ldots F(f_{2n}) d\mu = \text{E}(F(f_1)\ldots F(f_{2n}))
\]

where the sum is over all pairings of \( i_1,\ldots,i_{2n} \), i.e., \( i_1 < \ldots < i_{2n} \), and \( (i_1,i_2,\ldots,i_{2n}) \) is a permutation \( 1,\ldots,2n \).

\[
\Lambda^2(H) = L^2(H)^* + L^2(H) = L^2(H) + L^2(H)
\]

The space \( \Gamma(H) \) is intrinsic to the structure of \( H \) as a real Hilbert space. Thus if \( U: \Lambda^2 \to \Lambda^2 \) is an orthogonal mapping of one real Hilbert space into another, \( \Gamma(U) \) induces a unitary mapping \( \Gamma(U):\Gamma(H) \to \Gamma(K) \), where on \( \Gamma(H) \), \( U = \mathbb{E}(\Omega,A,u) \). Similarly if \( A: \Lambda^2 \to \Lambda^2 \) is an orthogonal mapping of one real Hilbert space into another, \( \Gamma(A) \) induces a unitary mapping \( \Gamma(A):\Gamma(H) \to \Gamma(K) \), where on \( \Gamma(H) \), \( A = \mathbb{E}(\Omega,A,u) \). Any contraction \( \Lambda^2 \to \Lambda^2 \) is an orthogonal mapping of one real Hilbert space into another, \( \Gamma(U) \) induces a unitary mapping \( \Gamma(U):\Gamma(H) \to \Gamma(K) \), where on \( \Gamma(H) \), \( U = \mathbb{E}(\Omega,A,u) \). Any doubly Markovian operation is a contraction from \( L^2 \) to \( L^2 \).

It turns out that if \( A \) 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 $L^2(\Omega,\mathcal{A},\mu)$ can be characterised as linear positivity preserving operators which are idempotent, of norm $\leq 1$ and preserve constants. We also know that for $p \geq 1, p \neq 2$, all linear operators $T$ on $L^p(\Omega,\mathcal{A},\mu)$, which are idempotent, contractive and such that $T1 = 1$ is necessarily a conditional expectation.

**Theorem 3.1 (Nelson Hypercontractivity Theorem).**

Let $A: \mathcal{H} \to \mathcal{K}$ be a contraction. Then $I(A)$ is a contraction from $L^q(\mathcal{H}) \to L^p(\mathcal{K})$ for $1 < q \leq \infty$ provided that

$$\|A\| < \left(\frac{2}{p-1}\right)^{1/2}$$

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


Acknowledgment

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