ON BUILDING NON-LINEAR DYNAMICAL STOCHASTIC PROCESS MODELS FROM EMPIRICAL DATA

by

Allen Jay Goldberg

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1986

ABSTRACT

Title of Dissertation: On Building Non-linear Dynamical Stochastic Process Models from Empirical Data

Allen Jay Goldberg, Doctor of Philosophy, 1986

Dissertation directed by: John S. Baras, Professor of Electrical Engineering, College of Engineering, University of Maryland

Wiener series is revisited as an approach to developing approximate models for arbitrary, scalar stochastic processes with continuous time parameter. The usual barrier to applying Wiener series to this problem is the difficulty of resolving the Wiener kernels from statistics of the given data. There are, however, known methods for determining Wiener kernels for Wiener series expansions of a white noise functional when one has access to both the functional output and the white noise excitation. One part of this paper is a unified discussion of Wiener series, Wiener series kernel identification procedures, and methods for approximating causal Wiener series by finite-dimensional causal bilinear systems driven by white Gaussian noise.

Attention is then directed toward using a construction due to Wiener and Nisio, built upon a typical path of the process data, which could serve as a starting candidate white noise functional. The given scalar process must be stationary, ergodic, and continuous in probability. The sequence of Wiener-Nisio functionals, when driven by the flow of the white noise excitation, gives rise to output processes which converge in finite-order distributions to the given process. The functional admits a causal Wiener series expansion which, when truncated, is realizable as a finite-order causal bilinear dynamical system driven by white noise excitation which also approximates the given process in finite distributions.

The paper then concludes with an analysis of this approach for two important classes of processes known to admit finite-order Wiener series expansions: Gaussian processes and Rayleigh power processes.

In Memory of My Mother,

Dorothy Feldman Goldberg

ACKNOWLEDGEMENT

I want to thank the many friends who have provided advice, encouragement and assistance during the writing of this paper. Special thanks go to my dear wife Bonnie, my children Emily and Sarah, my father and sister, and all the rest of my family for being so patient and understanding. I am extremely grateful to Dr. John Baras, my advisor, for his example and vision, and to all my professors from the University of Maryland, who helped me grow to new limits. I am endebted to my rare boss, Dr. Gerald Friedman, and my dear friend and colleague, Dr. Joseph P. Lawrence III, who provided steadfast encouragement and made countless sacrifices so that I could finish this endeavor. Many thanks to my co-workers at Naval Research Laboratory for setting such high standards and for making team research such a rewarding experience. I would also express my sincere appreciation to Dora Wilbanks, Judy Kogok, and the NRL publications staff for cheerfully expediting the type-setting of this paper.

Table of Contents

LIST OF FIGURES

"The causes of events are even more interesting than the events themselves."
...Cicero

1. Introduction

This paper is directed toward the development of theory and techniques for constructing dynamic stochastic models for general (non-Gaussian) scalar continuous-time stochastic processes directly from empirical data. The goal is to develop a realization algorithm for building a noise generator consisting of finite-order non-linear systems driven by white Gaussian noise. Whenever such a generator is exercised, it generates sample paths whose ensemble statistics (probability law) match that of the given data.

The role of such models is a means for synthesizing paths on demand which are not likely the same as given data, but which are statistically indistinguishable from given data. The mechanism for generating these paths need not reflect the true underlying physical mechanism of evolution of the given data. Nevertheless, it must be physically realizable and must produce statistically faithful sample paths. The realization procedure and the model framework should be capable of handling general processes in wide-ranging applications. In particular, the underlying process need not be Gaussian, and the application may require match-up of higher order moments than simply the mean or autocorrelation (or power spectrum). The model does not necessarily have to be of minimal order or dimension. However, a rigorous framework must be established to determine the benefit (or lack of benefit) of increased model order and complexity.

The approach taken is largely due to Wiener: that Wiener series, or the Homogeneous Chaos, serves as the fundamental approximation framework for the representative models. There are two principal thrusts to the paper: to clarify important interrelationships among various representations of Wiener series, their realization and identification, and to propose a candidate white noise functional, the Wiener-Nisio functional, which could serve as a basis from which a Wiener series could be built and then realized.

The organization of this paper is as follows. Sections 2 and 3 discuss the pair relationship of Wiener increments processes and white noise. Both excitations are defined carefully in conjunction with integrals and without resort to singularity functions, and this approach applies throughout the paper. Section 3 also introduces the concept of a functional of white noise which generates *processes* by allowing the white noise to flow. This is a useful setting for introducing Wiener series.

The next sections deal with various forms of Wiener series and their interrelationships. Section 5 reviews the properties of Fourier-Hermite Expansions. The Cameron-Martin theorem is described, particularly as it applies to finite mean-square functionals of white noise defined on the entire real line. The concept of a *lossy* process is introduced and discussed. Section 6 discusses multiple Wiener-Ito integrals. The material adopts Ito's approach and clarifies several details. The concept of avoiding the diagonal is emphasized both here and in other sections as being fundamental to the orthogonality properties of the Wiener series expansion. Section 7 discusses one form of white noise realization of Wiener-Ito integrals, the Wiener G-functionals. The discussion shows that, despite casual appearances, the G-functionals, like the Ito functionals, are homogeneous. Section 8 shows how the G-functionals can be realized as a telescoping series of Volterra operators driven by white noise and as stochastic differential equations. It further shows how G-functionals can be obtained from the lead Volterra operator directly without the need of additional states associated with the derived Volterra operators. Section 9 discusses two well-known methods of identifying Wiener kernels by input-output cross-correlation.

Section 10 explores the feasibility of using a sequence of Wiener-Nisio functionals as a candidates for building Wiener series models. The processes to be modelled must be stationary, ergodic, and continuous in probability. The Wiener-Nisio functionals are computed using a typical path of the data as the raw statistic. They are shown to converge in finite distributions to the given process, and this property is passed on to Wiener series realizations of the functional.

Section 11 then summarizes all that precedes in the form of an algorithm that might be used to realize general stochastic processes. Section 12 and 13 critically study the behavior of the resulting models for two important classes of processes, Gaussian and Rayleigh, respectively. In both of these examples, where simple finite series expressions are known to exist, the models resulting from this algorithm are shown to be problematical. A concluding section raises issues for further research.

"Things are always at their best in the beginning." ..... Pascal

2. Normal Random Measure and First-Order Wiener Integrals

The purpose of this section is to define the *first-order Wiener integral*:

$$\int f(t)\, dW(t) \tag{2.0.1}$$

of a *sure* or *deterministic* finite mean-square function $f(t) \in L_2(T)$ with respect to Wiener increments $dW(t)$ using the concept of Wiener random measure β_W. These integrals are fundamental to process representations; for example, they are the basis for the spectral representation of all stationary finite mean-square Gaussian processes [DOO, WON]. These integrals are also fundamental to all the nonlinear white noise functionals comprising the bulk of this paper. The development here of the Wiener integral follows the approach by Ito in [ITO1].

2.1 Wiener Random Measure and Wiener Increments Process

A *Gaussian process* $\{X_t,\ t \in T\}$ is a collection of random variables $X_t = X_t(\omega) = X(\omega,t) = X(t)$ such that every finite collection $X(t_1),\ldots,X(t_n)$, $t_1,\ldots,t_n \in T$, is multivariate Gaussian. In this paper, Gaussian processes always are assumed to have zero mean: $\mathbf{E}\{X(t)\} = 0$. Given any real-valued function $R_X(t,s)$ $<\infty$; $t,\ s \in T$, satisfying:

$$R_X(t,s) = R_X(s,t), \quad \text{and} \tag{2.1.2}$$

$$\Sigma\ c_i\ \overline{c}_j\ R(t_i,\ t_j) \geqslant 0; \text{ for any } t_1,\ \ldots,\ t_n \in T \text{ and any complex numbers } c_1,\ \ldots,\ c_n\ , \tag{2.1.3}$$

we can establish a sequence of finite-order Gaussian probability distributions for $\{X(t_1),\ldots,X(t_n)\}$, $n = 1,2,\ldots$, with $R_X\ (t_i,t_j)$, $i,j = 1,\ldots,n$ as the autocovariance matrix for each n, the finite-order distributions obeying a consistency condition as lower order distributions are obtained from higher order distributions by "integrating out" the excess arguments. These distributions, in turn, by Kolmogoroff's existence theorem [BIL1], uniquely determine a probability measure $\mathbf{P}(\omega)$ for the process X_t defined on the space of continuous functions (sample paths) on R^T. The function

$$R_X(t,s) = \mathbf{E}\{X_t, X_s\} = \int X_t(\omega) X_s(\omega)\, d\,\mathbf{P}(\omega) \tag{2.1.4}$$

is then, of course, the autocorrelation function for the process X_t.

Let $(T,\mathbf{B},m)$ be a measure space, and let $\mathbf{B}^* = \{E: E \in \mathbf{B}, m(E) < \infty\}$. A Gaussian process $\beta(E,\omega)$ (defined on sets E as opposed to point values t) is called a *normal random measure* on $(T,\mathbf{B},m)$ if

$$\mathbf{E}\{\beta(E)\,\beta(E')\} = m(E \cap E') \text{ for any } E,E' \in \mathbf{B}^*. \tag{2.1.5}$$

Since $m(E \cap E') = m(E' \cap E)$, and $\Sigma\, c_i\, \bar{c}_j m(E_i \cap E_j) = \int |\sum_i c_i\, \chi_i(t)|^2\, m(dt) \geqslant 0$, $\chi_i(t)$ being the indicator of set E_i, the process β is well defined, with a uniquely defined probability measure, for any measure space $(T,\mathbf{B},m)$.

Let $\beta(E)$ be a normal random measure on $(T,\mathbf{B},m)$. If $E_1,E_2,\ldots$ are disjoint, then $\beta(E_1), \beta(E_2),\ldots$ are independent. Furthermore, if $E = E_1 + E_2 + \ldots \in \mathbf{B}^*$, then

$$\beta(E) = \Sigma\, \beta\,(E_n), \tag{2.1.6}$$

converging in the mean-square sense. Since $\beta(E_1), \beta(E_2),\ldots$ are independent, this then implies almost sure convergence by Levy's theorem [CHU, p. 120]. This result illustrates how β may be regarded as a random measure.

With primary interest on continuous-parameters process models and stochastic differential equations, we further restrict our interest to measures m which have the following continuity property: For any $E \in \mathbf{B}^*$ and $\epsilon > 0$, there exists a decomposition of E:

$$E = \sum_{i=1}^{n} E_i \tag{2.1.7}$$

such that

$$m(E_i) < \epsilon,\ i = 1, \ldots, n. \tag{2.1.8}$$

We now specialize this for Wiener processes. Several references [WIE2, SHU, MCK3] describe, in varying degrees of detail and sophistication, the construction and properties of W_t, the *Wiener process*, a zero mean Gaussian process with (almost surely) continuous sample paths usually defined on $[0,\infty)$ and with autocorrelation function $R_W(t,s) = \min(t,s)$. The definition of Wiener process can be extended to $(-\infty, \infty)$ [WON] by piecing together two Wiener processes $W_{1,t}$ and $W_{2,-t}$ using the modified autocorrelation

$$R_W\ (t,s) = (1/2)(t| + |s| - |t-s|). \tag{2.1.9}$$

The adoption of the origin as the reference time $t=0$ is purely arbitrary; our ultimate interest is in the increments process dW_t, which does not depend on any particular choice of reference time. For example, while the Wiener process W_t is not stationary, with autocorrelation function (2.1.9) depending on the choice of reference time, the increments process dW_t is stationary (see (2.1.11) and (2.1.5)), and its statistics do not depend on the choice of reference time at all.

The associated probability measure on the space of continuous functions is the *Wiener probability measure.* We construct the *Wiener random measure* β_W from the

Wiener process W_t, first for intervals: $\beta_W([t,t+dt]) = W_{t+dt} - W_t$; then, by extension in the mean-square sense (and almost surely), to measurable sets on $(-\infty,\infty)$, as described above. What results is a normal random measure corresponding to measure $m =$ the Lebesgue measure on $T = (-\infty,\infty)$. We can then associate the customary notion of the Wiener process with:

$$W_t = W(t) = \beta_W([0,t]), \tag{2.1.10}$$

and the customary notion of the *Wiener increments process* with:

$$dW_t = dW(t) = \beta_W([t,t+dt]). \tag{2.1.11}$$

The Wiener probability measure, which underlies the Wiener process W_t, also underlies both β_W and dW_t, and frequently, the distinction between them may be overlooked.

2.2 First-order Wiener Integral

We now define the *first-order Wiener integral* of a *sure* finite mean-square function $f(t) \in L_2(T)$ as follows. If $f(t) = \chi_E(t)$, an indicator function, then

$$I_1(f) = I_1(\chi_E) = \int \chi_E(t)\ dW(t) = \beta_W(E). \tag{2.1.12}$$

We also note that

$$\mathbf{E}\{I_1(\chi_E)\ I_1(\chi_F)\} = m(E \cap F), \tag{2.2.13}$$

so that $I_1(\chi_E)$, $I_1(\chi_F)$, being Gaussian, are independent if E and F are non-overlapping. (This is more relaxed than requiring them to be disjoint. They can, for example, share a common end point.)

If $f(t)$ is an elementary function:

$$f(t) = \sum_{i=1}^{n} a_i\, \chi_i, \tag{2.2.14}$$

χ_i the indicator of set E_i, where we always choose the sets E_i to be disjoint, then

$$I_1(f) = \sum_{i=1}^{n} a_i \, I_1(\chi_i). \tag{2.2.15}$$

Finally, for $f(t) \in L_2(T)$, let $f_i(t)$ be a sequence of elementary functions (2.214) converging in the mean-square sense to $f(t)$. Then

$$I_1(f) = \lim_{n \to \infty} \sum_{i=1}^{n} I_1(f_i), \tag{2.2.16}$$

where the convergence is in the mean-square sense, and therefore, by virtue of the independence of the $I_1(f_i)$, is also almost sure.

"Energy is eternal delight." .... W. Blake

3. White Noise

We now proceed to define Gaussian white noise, which we will call simply white noise, and to show its relationship to the Wiener process. White noise is a conceptual model for random thermal excitation, such as the force that buffets particles about in Brownian Motion (for which the Wiener process is a model). It represents an idealization of such an excitation in that the amount of constant thermal energy per unit bandwidth available is limited only by the bandwidth of the physical system experiencing its effects. It is what we conclude must be the excitation at a linear dynamical system's input to give rise to Gaussian stochastic process outputs with spectrum matching the power gain per unit bandwidth of the system. For any one physical system with finite bandwidth, there is an equivalence class of distinct signal processes which are indistinguishable from one another that can be regarded as the white noise, which have smooth sample paths; white noise is the limiting case that applies for virtually all finite-bandwidth systems simultaneously.

Mathematically, white noise belongs to a class of generalized functions which, like the Dirac delta function, only takes on meaning in conjunction with integrals and the integral properties. Like the delta function, white noise may be envisioned as a limit of a sequence (in this case, of Gaussian processes) of ever-widening flat power spectrum.

White noise bears a close relationship to the Wiener increments process which permits interchange of first-order white noise integrals with first-order Wiener integrals. This interchange is a useful technique for studying non-linear finite mean-square functionals of white noise, and the behavior of associated nonlinear systems to white noise excitation, from several important points of view.

We begin by defining the integral

$$\int_T f(t)\ X(t)\ dt \tag{3.0.1}$$

of the product of an $L_2(T)$ sure function $f(t)$ with a second-order (Gaussian) process $X(t)$ as a prelude to defining the white noise integral

$$\int_T f(t)\ n(t)\ dt. \tag{3.0.2}$$

This contrasts with the Wiener integral

$$\int_T f(t)\ dW(t) \tag{3.0.3}$$

developed in the previous section.

3.1 Definition: Quadratic-Mean (Q.M.) Stochastic Integral [WON]

Let $X(t)$ be a second-order stochatic process defined on $T = [a,b]$, $R_X(t,s) = \mathrm{E}\{X(t)X(s)\}$, and let $f(t)$ be a real-valued function defined on T. Let π_n be a partition of T:

$$\pi_n = \{a = t_{n,0} < t_{n,1} < \ \ldots < t_{n,n} = b\} \tag{3.1.1}$$

$$\delta_n = \max_{[1 \leqslant i \leqslant n]} (t_{n,i} - t_{n,i-1}) \tag{3.1.2}$$

such that

$$\delta_n \to 0 \text{ as } n \to \infty. \tag{3.1.3}$$

Then the *q.m.* integral

$$\int_T f(t)\ X(t)\ dt = \lim q.m. \sum_{i=1}^{n-1} f(s_{n,i})\ X(s_{n,i})\ (t_{n,i+1} - t_{n,i})\,, \tag{3.1.4}$$

where

$$t_{n,i} \leqslant s_{n,i} \leqslant t_{n,i+1}. \tag{3.1.5}$$

The integral is well defined provided that the $q.m.$ limit exists and is independent of the choice of partitions $\{\pi_n\}$ and for each π_n the choice of $\{t_{n,i}\}$.

If $T = (-\infty,\infty)$ the $q.m.$ integral is defined as:

$$\int_{-\infty}^{\infty} f(t)\ X(t)\ dt = \lim_{\substack{a \to -\infty \\ b \to \infty}} q.m. \int_{[a,b]} f(t)\ X(t)\ dt. \tag{3.1.6}$$

Remark:

The integral exists if and only if $\int_T \int_T f(t) f(s)\ R_X\ (t,s) dt\ ds$ exists as a Riemann integral. We could also choose to define $\int_T f(t) X(t) dt$ as a Lebesque integral of a sample function $X(t,\omega)$; existence of such an integral is ensured if the process is measurable and if $\int_T |f(t)| \mathbf{E} |X(t)| dt < \infty$. (See discussion of normal random measure.)

3.2 Definition: (Gaussian) White Noise. [WON]

A sequence of quadratic mena-square continuous Gaussian processes $X_n(t), t \in T$, is said to converge to white noise $n(t)$, denoted $X_n(t) \to n(t)$, if

a) for each $f \in L_2(T)$, the sequence of $q.m.$ integrals

$$Q_n(f) = \int_T f(t)\ X_n(t)\ dt \to Q(f) \tag{3.2.1}$$

is $q.m.$ convergent, and

b) there exists a positive constant S_0 such that, for every $f,g \in L_2(T)$,

$$\lim_n \mathbf{E}\ \{Q_n(f)\ Q_n(g)\} = S_0 \int_T f(t)\ g(t) dt. \tag{3.2.2}$$

Remark:

An alternate (and popular!) method for expressing (3.2.1) and (3.2.2) is:

$$\begin{aligned} \mathbf{E}\{Q(f)\,Q(g)\} &= \int_T \int_T [\mathbf{E}\{n(t)\,n(s)\}]\, f(t)\, g(s)\, dt\, ds \\ &= \int_T \int_T S_0\, \delta(t-s)\, f(t)\, g(s)\, dt\, ds \\ &= S_0 \int_T f(t)\, g(t)\, dt. \end{aligned} \tag{3.2.3}$$

The use of delta functions in the analysis of non-linear systems must be tempered. In certain applications, such as computing high-order moments of Gaussian processes, delta functions can be used effectively to simplify bookkeeping. On the other hand, I strongly discourage the exclusive use of (3.2.3) to *define* white noise. It is too incomplete a description to shed light on important properties of white noise non-linear functionals, many of which are discussed at length in this paper. On the strength of the delta function approach alone, it is virtually impossible to understand in what sense the Wiener G-functionals of white noise behave like homogenous operators, why the Lee-Schetzen identification method fails to identify G-functional kernels on the diagonals, and why Ito stochastic differential equations need the Wong-Zakai correction to produce accurate results.

The next theorem ties together first-order white noise and first-order Wiener integrals.

3.3 Theorem: Altenate Representation of First-Order White Noise Integral. [WON]

Let $X_n(t) \rightarrow n(t)$, let $f(t) \in L_2(T)$, and let

$$Q(f) = \int_T f(t) n(t)\, dt = \lim\ q.m. \int_T f(t)\, X_n(t)\, dt. \tag{3.3.1}$$

Then,

$$Q(f) = S_0^{-\frac{1}{2}} \int_T f(t)\, dW(t), \tag{3.3.2}$$

a Wiener integral, with $\{W(t), t \in T\}$ the Wiener process determined by

$$W(t) = S_0^{-\frac{1}{2}} \lim q.m. \int_0^t X_n(s)\, ds. \tag{3.3.3}$$

Proof:

Define $W(t)$ by (3.3.3), and let χ_{ab} be the indicator of $[a,b]$. Then

$$W(b) - W(a) = S_0^{-\frac{1}{2}} \lim q.m. \; Q_n\,(\chi_{ab}). \tag{3.3.4}$$

From (3.2.2),

$$\mathbf{E}\,\{W(b) - W(a))\,(W(d) - W(c))\} = \int_T \chi_{ab}(t)\, \chi_{cd}(t)\, dt, \tag{3.3.5}$$

hence, $W(t)$ has orthogonal increments $dW(t)$ and

$$\mathbf{E}\,\{dW(t)\; dW(s)\} = \delta_{t,s}\; dt. \tag{3.3.6}$$

Since $X_n(t)$ are Gaussian, $W(t)$ is also Gaussian, and is therefore the Wiener process. Let $\int_T f(t) dW(t)$ be defined as a Wiener integral for $f(t) \in L_2(T)$. If f is a step function, then easily by (3.3.4), $Q(f) = S_0^{\frac{1}{2}} \int_T f(t) dW(t)$. If f is not a step, let f_n be a sequence of step functions $f_n \rightarrow f$ in the L_2 sense. Then,

$$\begin{aligned} Q(f) &= Q(f_n) + Q(f-f_n) = S_0^{\frac{1}{2}} \int_T f_n(t)\; dW(t) + Q(f-f_n) \\ &\rightarrow S_0^{\frac{1}{2}} \int_T f(t)\; dW(t) + Q(0) = S_0^{\frac{1}{2}} \int_T f(t)\; dW(t). \end{aligned} \tag{3.3.7}$$

Q.E.D.

Remark:

This theorem provides the fundamental means for interchanging linear functionals of white noise with linear functionals of Wiener process increments. The relationship

$$\int_T f(t)\; n(t)\; dt = S_0^{\frac{1}{2}} \int_T f(t)\; dW(t) \tag{3.3.8}$$

is frequently expressed as

$$n(t)\ dt = S_0^{\frac{1}{2}}\ dW(t), \tag{3.3.9}$$

which suggests white noise is somehow the derivative of the Wiener process. Caution is advised in this interpretation, since the sample paths of the Wiener process (almost surely) are continuous everywhere, but nowhere differentiable! Equation (3.3.9) symbolically describes the equivalence of *time averages* of white noise to Wiener increments over distinct, *smoothing* intervals. Here again is the notion of an *integral*. Equation (3.3.9) is an elementary form of stochastic differential equation, a symbolic equation that actually stands for an *integral* equation, here (3.3.8).

For multiple integrals, interchange between white noise and Wiener increments integrals is more complex than (3.3.9), and in particular, depends on the adopted definition for the integral. This will be amplified in later sections.

3.4 Stationarity, Ergodicity and the Translation Group

In the development of the Fourier-Hermite series, the multiple Wiener integral and the Wiener-Nisio functional is the sections that follow, the time evolution of response of a system to white noise excitation is modeled in a particular way. First, the response of the system, $Y(t)$, is evaluated at a particular time, say $t = 0$. This value is a *functional* F of the entire white noise path $n(\cdot)$. In the case of Wiener expansions, for example, the functional is the infinite sum of multiple Wiener integrals. The response *process* $Y(t)$ is then obtained, either by letting the excitation $n(\cdot)$ *shift* or *flow*, or by altering the functional (for example, by changing the integral kernels), as follows:

If

$$Y(0) = F(n(\cdot),0),$$

then

$$Y(t) = F(n(\cdot+t),0) = F(n(\cdot),t). \tag{3.4.1}$$

The flow of the white noise $n(t)$ can be regarded as the action of a *translation group operator* T_t on outcomes in the probability space on which $n(t)$ is built. (See [DOO] or [WON].) The *stationarity* of $n(t)$ is inhereted from the *metric transitivity* or invariance of the probability measure of the outcomes under T_t.

An outcome that is almost surely preserved under translation is called an *invariant* random variable. If every invariant random variable associated with a stationary process is almost surely a constant, then the process is called *ergodic*. If X_t is a separable measurable stationary ergodic process, and f is any Borel function such that $\mathbf{E}\{|f(X_0)| < \infty$, then

$$\lim_{a \to \infty} (1/2a) \int_{-a}^{a} f(X_t)\, dt = \mathbf{E}\,\{f(X_0)\}\ a.s. \tag{3.4.2}$$

The processes of interest here, modeled by the flow of the white noise acting on a functional, such as the integral functionals in (3.3.8), inheret the stationary, ergodic properties of the white noise itself.

An explicit example of the translation operator T_t can be found in the *spectral representation* [DOO, WON] of a stationary Gaussian process X_t via the Wiener integral:

$$X_t = \int_{-\infty}^{\infty} e^{i2\pi\lambda t} dW_\lambda, \tag{3.4.3}$$

where the (generalized) Wiener process W_λ defined on frequency parameter λ has independent increments satisfying:

$$E\{dW_\lambda^2\} = d\sigma(\lambda), \tag{3.4.4}$$

where $\sigma(\lambda)$ is the (cumulative) power spectral distribution of the process. The translation operator T_t for X_t presents itself in the form of the familiar linear frequency-dependent phase shift $e^{j2\pi\lambda t}$ associated with each time t.

"Adde parvum parvo magnus acervus erit"
[Add little to little and there will be a great heap]
. . . . Ovid

5. Fourier-Hermite Expansions

The celebrated Cameron-Martin Theorem serves an essential role in this paper. The Cameron-Martin theorem proves that every finite mean-square functional of a Wiener increments process on a compact interval can be approximated in the mean-square sense by an orthogonal expansion of products of Hermite polynomials acting upon orthogonal linear functionals of the Wiener process increments. The resulting Grad-Barrett-Hermite expansion of a finite mean-square functional [CAM, BAR 1-2], and provides a conceptual block diagram for all such functionals. It leads immediately to Fourier-Hermite expansions in terms of products of Hermite polynomials acting upon orthogonal linear functionals of white noise, where the white noise is always smoothed by linear filtering prior to undertaking nonlinear transformations. The Cameron-Martin theorem is also crucial in proving the completeness of the multiple Ito integrals and the Wiener G-functionals.

This section begins with some reference formulas for moments of Gaussian random variables and properties of Hermite polynomials expressed in several, popular forms. The Cameron-Martin Theorem is then proven, followed by some discussion of the nature of the expansion when the functional to be approximated is causal. Some remarks are then made on adapting the Fourier-Hermite expansions to represent causal functionals defined on $T = [0,\infty)$ using the orthonomal Laguerre functions as the kernels of the linear filters. The resulting expansion describes a "*nonexplosive* [finite mean-square under white-noise excitation] time-invariant nonlinear system with *noninfinite memory.*" [SHE1, also WIE2, RUG1] This mirrors the observation by McKean [MCK1] that, if a finite mean-square process defined on the real line admits a Wiener expansion representation, it must satisfy a *mixing* property [BIL1].

5.1 Theorem: Averages of Products of Gaussian Random Variables [SHE1, ROO]

Let $X_1, \ldots, X_n$ be jointly Gaussian, $E\{X_i\} = 0$. $E\{X_i^2\} = 1$, $i = 1, \ldots, n$ (normalized). Then

$$E\{X_i \ldots X_n\} = 0, \qquad n = 2N + 1 \text{ (odd)}, \tag{5.1.1}$$

$$= \Sigma\Pi R_{ij}, \quad n = 2N \text{ (even)}, \tag{5.1.2}$$

where $\Sigma\Pi$ denotes the sum of N-products of moments $R_{ij} = E\{X_i X_j\}$ which represent the *distinct* ways of partitioning the indices $1, \ldots, 2N$ into N pairs. The number of terms in $\Sigma\Pi$ is:

$$\frac{(2N)!}{N!2^N}. \tag{5.1.3}$$

Example ($n = 4, N = 2$):

$$E\{X_1 X_2 X_3 X_4\} = R_{12} R_{34} + R_{13} R_{24} + R_{14} R_{23}. \tag{5.1.4}$$

Proof: Let

$$\Gamma_X(a_i, \ldots, a_n) = E\{e^{j(a_1 X_1 + \ldots + a_n X_n)}\} \tag{5.1.5}$$

be the moment generating function associated with $X_1, \ldots, X_n$. First, we expand Γ_X by Taylor series:

$$\Gamma_X(a_1, \ldots, a_n) = \sum_{k_1 = 0}^{\infty} \sum_{k_n = 0}^{\infty} c_{k_1 \cdots k_n} a_1^{k_1} \ldots a_n^{k_n} \tag{5.1.6}$$

where

$$c_{k_1 \cdots k_n} = \frac{1}{k_1! \ldots k_n!} \frac{\partial^{k_1}}{\partial_{a_1}^{k_1}} \cdots \frac{\partial^{k_n}}{\partial_{a_n}^{k_n}} \Gamma_x(a_1, \ldots, a_n)\Bigg|_{a_1 = \cdots = a_n = 0} \tag{5.1.7}$$

$$= \frac{1}{k_1! \cdots k_n!} j^{k_1} \ldots j^{k_n} E\{X_1^{k_1} \cdots X_n^{k_n}\}. \tag{5.1.8}$$

Thus,

$$\Gamma_X(a_1,\ldots,a_n) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} E\{X_1^{k_1} \ldots X_n^{k_n}\} \frac{(ja_1)^{k_1} \cdots (ja_n)^{k_n}}{k_1! \ldots k_n!}. \quad (5.1.9)$$

Now, the only term containing $E\{X_1 \ldots X_n\}$ is associated with $k_1 = \ldots = k_n = 1$:

$$a_1 \ldots a_n (j)^n \; E\{X_1 \ldots X_n\}. \quad (5.1.10)$$

We set this aside momentarily, and expand Γ_X another way. Since $X_1 \ldots X_n$ have zero mean, unit variance, and are jointly Gaussian,

$$\Gamma_X(a_1,\ldots a_n) = e^{-\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j R_{ij}\right)} \quad (5.1.11)$$

is Gaussian and is expressed as shown. Using

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad (5.1.12)$$

$$\Gamma_X(a_1,\ldots,a_n) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(-\frac{1}{2}\right)^k \left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j\, R_{ij}\right]^k \quad (5.1.13)$$

$$= 1 + \left(-\frac{1}{2}\right) \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} a_{k_1} a_{k_2} R_{k_1 k_2}$$

$$+ \frac{1}{2!}\left(-\frac{1}{2}\right)^2 \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \sum_{k_3=1}^{n} \sum_{k_4=1}^{n} a_{k_1} a_{k_2} a_{k_3} a_{k_4} R_{k_1 k_2} R_{k_3 k_4} \quad (5.1.14)$$

$$+ \ldots.$$

Each set of terms in (5.1.14) has an even number of factors of the a_i's. Therefore, we will find a non-zero match for (5.1.10) only if n is even. Thus,

$$E\{X_1 \ldots X_n\} = 0,\; n = 2N + 1 \text{ (odd.)} \quad (5.1.15)$$

as hypothesized.

We now consider the case $n = 2N$ (even). The only terms that can contain products of the form $a_{k_1} \ldots . a_{k_n}$ are those for which $k = n$:

$$\frac{1}{N!}\left(-\frac{1}{2}\right)^N \sum_{k_1=1}^{2N} \cdots \sum_{k_{2N}=1}^{2N} R_{k_1 k_2} \ldots R_{k_{2N-1} k_{2N}}\, a_{k_1} \ldots a_{k_{2N}}. \quad (5.1.16)$$

Since, in (5.1.10), the product $a_1 a_2 \ldots a_n$ has all unique subscript values, the matching terms from (5.1.16) must have unique subscript values as well:

$$\frac{1}{N!}\left(-\frac{1}{2}\right)^N a_1 \ldots a_{2N} \, \Sigma_{k_i\text{'s}_{\text{distinct}}} R_{k_1 k_2} \ldots R_{k_{2N-1}\, k_{2N}} \tag{5.1.17}$$

Since $R_{k_i\, k_j} = R_{k_j k_i}$, and there as N pairs of factors, so there are 2^N terms of the sum which are identical by virtue of commutation symmetry. In addition, we note that $R_{k_1 k_2} R_{k_3 k_4} = R_{k_3 k_4} R_{k_1 k_2}$. There are $N!$ terms (permutations) that are equivalent by virtue of this symmetry. Thus, (5.1.17) reduces to:

$$(-1)^N \, (a_1 a_2 \ldots a_{2N}) \, \Sigma \, \Pi \, R_{ij} \tag{5.1.18}$$

as described under (5.1.2). The result (5.1.2) follows by equating (5.1.10) and (5.1.18)

Q.E.D.

5.2 Hermite Polynomials and Hermite Functions

The Hermite *polynomials* are a complete orthogonal set with respect to Gaussian measure. Table 5.2.1 summarizes the properties of the Hermite polynomials as they appears in random forms in several important papers and books pertaining to nonlinear Wiener functionals.

Beyond this, the Hermite *functions*

$$k_n(x) = H_n(x) \sqrt{\frac{p(x)}{q_n}} \tag{5.2.2}$$

are complete orthornormal (C.O.N.) set on $L_2(-\infty,\infty)$. [DYM] This means that the integrals

$$I_1\,(k_n,n) = \int_{-\infty}^{\infty} k_n(t)\; n(t)\,dt \tag{5.2.3}$$

Table 5.2.1a — Properties of Hermite Polynomials [ABR, pp. 771-802]

Symbolic Representation:	$H_n(ax)$
Weighting function:	$p(x) = e^{-a^2 x^2}$
Orthogonality and normalization:	$\int_{-\infty}^{\infty} H_m(ax)\, H_n(ax)\, p(x)\,dx = \begin{cases} 0, & m \neq n \\ q_n = \dfrac{2^n n!}{a}\sqrt{\pi}, & m = n \end{cases}$
Recursion:	$H_{n+1}(x) = 2xH_n(x) - 2n\, H_{n-1}(x)$
Polynomial Expansion:	$H_n(x) = \sum_{i=0}^{\left[\frac{n}{2}\right]} \dfrac{(-1)^i n!}{i!(n-2i)!} (2x)^{n-2i}$

First Few Polynomials:

$n=0$	$H_n(x) = 1$
$n=1$	$2x$
$n=2$	$4x^2 - 2$
$n=3$	$8x^3 - 12x$
$n=4$	$16x^4 - 48x^2 + 12$
$n=5$	$32x^5 - 160x^3 + 120x$
$n=6$	$64x^6 - 480x^4 + 720x^2 - 120$
$n=7$	$128x^7 - 1344x^5 + 3360x^3 - 1680x$

Table 5.2.1b — Properties of Hermite Polynomials [IT01]

Symbolic Representation: $H_n\left(\frac{x}{\sqrt{2}}\right)$ (see 5.2.1a)

Weighting function: $p(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

Orthogonally and normalization:

$$\int_{-\infty}^{\infty} H_m\left(\frac{x}{\sqrt{2}}\right) H_n\left(\frac{x}{\sqrt{2}}\right) p(x)\,dx = \begin{cases} 0, & m \neq n \\ q_n = 2^n n!, & n = n \end{cases}$$

Table 5.2.1c — Properties of Hermite Polynomials [CAM]

Symbolic Representation: $H_n^c(x) = \frac{1}{\sqrt{2^n n!}} H_n(x)$

Weighting function: $p(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$

Orthogonality and normalization:

$$\int_{-\infty}^{\infty} H_m^c(x)\, H_n^c(x) p(x)\,dx = \begin{cases} 0, & m \neq n \\ q_n = 1, & m = n \end{cases}$$

Table 5.2.1d — Properties of Hermite Polynomials

(Section 5.3, modified version of [CAM])

Symbolic Representation: $H_n^c\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{\sqrt{2^n n!}} H_n\left(\frac{x}{\sqrt{2}}\right)$

Weighting function: $p(x) = \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2}}{2}$

Orthogonality and normalization:

$$\int_{-\infty}^{\infty} H_m^c\left(\frac{x}{\sqrt{2}}\right) H_n^c\left(\frac{x}{\sqrt{x}}\right) p(x)\,dx = \begin{cases} 0, & m \neq n \\ q_n = 1, & m = n \end{cases}$$

Table 5.2.1e — Properties of Hermite Polynomials [SHE1, pp. 405-411]

Symbolic Representation: $\tilde{H}_n(x) = (\sqrt{S_o/2})^n \, H_n \left(\frac{x}{\sqrt{2S_o}} \right)$

Weighting function: $p(x) = \frac{1}{\sqrt{2\pi S_o}} e^{\frac{-x^2}{2S_o}}$

Orthogonality and normalization:

$$\int_{-\infty}^{\infty} \tilde{H}_m(x)\, \tilde{H}_n(x)\, p(x)\, dx = \begin{cases} 0, & m \neq n \\ q_n = S_o^n n!, & m = n \end{cases}$$

Recursion:

$$\tilde{H}_{n+1}(x) = x\, \tilde{H}_n(x) - S_o\, \tilde{H}'_n(x)$$
$$\tilde{H}_{n+1}(x) = x\, \tilde{H}_n(x) - S_o\, \tilde{H}_{n-1}(x)$$

Polynomial Expansion:

$$\tilde{H}_n(x) = \sum_{i=0}^{\left[\frac{n}{2}\right]} \frac{(-1)^i n!}{i!(n-2i)!} \left(\frac{S_o}{2} \right)^i x^{n-2i}$$

First Few Polynomials:

$n = 0$	$\tilde{H}_n(x) = 1$
$n = 1$	x
$n = 2$	$x^2 - S_o$
$n = 3$	$x^3 - 3\, S_o x$
$n = 4$	$x^4 - 6\, S_o\, x^2 + 3S_o^2$
$n = 5$	$x^5 - 10\, S_o\, x^3 + 15S_o^2 x$
$n = 6$	$x^6 - 15S_o\, x^4 + 45S_o^2 x^2 - 15S_o^3$
$n = 7$	$x^7 - 21S_o\, x^5 + 105S_o^2 x^3 - 105S_o^3 x$

form a C.O.N. span of the *non-causal linear* finite near-square functionals of white noise $n(t)$. Section 5.4 discusses the Laguerre functions, a C.O.N. set on $L_2(0,\infty)$ which in turn lead to a C.O.N. span of the *causal linear* finite mean square functionals of white noise.

A further property of Hermite functions sheds valuable light on the geometry of $L_2(-\infty,\infty)$ and the Fourier integral. The scaled Hermite functions

$$\tilde{K}_n(x) = K_n(\sqrt{2\pi}x) \tag{5.2.4}$$

are *eigenfunctions* of the Fourier transform of $f \in L_2(-\infty,\infty)$:

$$F(f)(\gamma) = \int_{-\infty}^{\infty} f(x)\, e^{-j2\pi\gamma x} dx \tag{5.2.5}$$

with corresponding eigenvalues $(-j)^n$:

$$F(\tilde{K}_n)(\gamma) = (-j)^n\, \tilde{K}_n(\gamma). \tag{5.2.6}$$

Letting

$$< f, \tilde{K}_n > = \int_{-\infty}^{\infty} f(x)\, \tilde{K}_n(x) dx \tag{5.2.7}$$

we can therefore write the Fourier transform

$$F(f)(\gamma) = \sum_{n=0}^{\infty} < f, \tilde{K}_n > \tilde{K}_n(\gamma) \tag{5.2.8}$$

as a *countable series expansion.* [DYM] .

This property (countable series expansion of the Fourier integral) may seem surprising if one views the Fourier transform as the extension of Fourier series, with countable sinusoids, to the case of uncountable sinusoids. The rapidly-decreasing character of every $f \in L_2(-\infty,\infty)$ renders uncertain the precise identities of the sinusoids. These individual identities could only be resolved by perfectly coherent filtering. Just as one cannot resolve the reals from the rationals without "waiting forever," so the continuum of exponentials cannot be resolved beyond countability by test functions $f \in L_2(-\infty,\infty)$.

5.3 Theorem: Cameron-Martin [CAM]

Let $F[W(\cdot)]$ denote a real or complex valued functional defined on the standard Wiener process $W(t)$, $t \in T = [a,b]$ (an interval), and let $F[W(\cdot)]$ be finite mean-square with respect to Wiener measure:

$$\mathrm{E}\{\,|\,F\,[W]^2\,|\,\} = \int |\,F[W(\omega,\cdot)]|^2 d\mathbf{P}_W(\omega) < \infty \tag{5.3.1}$$

Let $H_n^N(u) = H_n^C(\frac{u}{\sqrt{2}})$ denote the particular, normalized version of n-th order Hermite polynominal described in table 5.2.1c-d, and let $\{\alpha_p(t)\}$, $p=1,2,\ldots$ be any complete orthonormal (C.O.N.) set of real functions belonging to L_2 (T).

Let

$$\Phi_{m,p}(W) = H_m^N\,[\,\int_T \alpha_p(t)dW(t)\,];\ m = 0,1,2,\ldots; \quad p = 1,2,\ldots, \tag{5.3.2}$$

and let

$$\Psi_{m_1\ldots m_p}(W) = \Phi_{m_1,1}(W)\ldots\Phi_{m_p,p}(W) \tag{5.3.3}$$

be the so-called Fourier-Hermite set. (Note that since $H_O^N \equiv 1$, we can freely append additional zero subscripts to Ψ; that is, $\Psi_{m_1\ldots mp0\ldots 0}(W) = \Psi_{m_1\ldots mp}(W)$.)

Then,

$$\mathrm{E}\{\,|\,F(W) - \sum_{m_1,\ldots m_N=O}^{N} A_{m_1\ldots m_N}\Psi_{m_1\ldots m_N}(W)\,|^2\,\} \to 0 \quad as\ N\to\infty \tag{5.3.4}$$

where

$$A_{m_1\ldots m_N} = \mathrm{E}\{F[W]\ \Psi_{m_1\ldots M_N}(W)\,\} \tag{5.3.5}$$

is the Fourier-Hermite coefficient associated with $\Psi_{m_1\ldots m_N}$.

Remark: In [CAM], the expansion is developed for functionals with respect to non-standard Wiener Process $X(t)$ which is related to $W(t)$ by

$$dX(t) = \frac{1}{\sqrt{2}} dW(t) \tag{5.3.6}$$

This makes all the associated linear functionals in the original proof have variance = $\frac{1}{2}$ rather than 1, and requires a particular form of Hermite polynominal $(H_n^C(u))$ and weight $p(u)$ to achieve the required normalizations. In this proof, normalization has been achieved with respect to the standard Wiener process $W(t)$, and so it differs slightly from the proof in [CAM].

5.3a Step 1: The Fourier-Hermite set is orthonormal.

The set $\{V_p\}$:

$$V_p = \int_T \alpha_p(t) dW(t) \tag{5.3.7}$$

of Gaussian random variables is independent, identically distributed, zero mean, unit variance. Therefore

$$E\{G(V_1, , V_p)\} = (2\pi)^{-\frac{p}{2}} \int_{-\infty}^{\infty} \cdots \int_{\infty}^{\infty} G(u_1, u_p) e^{\frac{-u_1^2}{2}} e^{\frac{-u_p^2}{2}} du_1 \ldots du_p \tag{5.3.8}$$

holds for every function G which is Lesbegue integrable with respect to Gaussian measure (r.h.s. of 5.38). Thus,

$$\begin{aligned} E\{\Psi_{m_1 \cdots m_n}(W)\Psi_{j_1 \cdots j_n}(W)\} \\ = (2\pi)^{-\frac{p}{2}} \int_{\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_{M_1}^N(j_1)\, H_{j_1}^N(u_n) \frac{e^{-u_1^2}}{2} \\ \ldots H_{m_n}^N(u_n)\, H_{j_n}^N(u_n) \frac{e^{-u_n^2}}{2} du_1 \ldots du_n . \end{aligned} \tag{5.3.9}$$

By virtue of the orthonormality of $H_i^N(u)$:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_i^N(u) H_j^N(u) e^{\frac{-u^2}{2}} du = \delta_{ij}, \tag{5.3.10}$$

we get

$$\mathbf{E}\{\Psi_{m_1 \cdots m_n}(W)\Psi_{j_1 \cdots j_n}(W)\} = \delta_{ij}. \tag{5.3.11}$$

5.3b Step 2: Best Approximation and Bessel Inequality

The mean-square error:

$$\begin{aligned} \mathbf{E}\{| F[w] - \sum_{m_1, \ldots, m_n = 0}^{N} B_{m_1 \cdots m_n} \Psi_{m_1 \cdots m_n}[W] |^2\} \\ = \mathbf{E}\,\{F[W]\,|^2\}. - 2Re. \sum A\bar{B} + \sum | B |^2 \\ = \mathbf{E}\{\,| F[W] |^2\} - \sum | A |^2 + \sum | A - B |^2 \end{aligned} \tag{5.3.12}$$

follows from (5.3.3), (5.3.5) and (5.3.11) and Lemma 5.3a. The error is minimized by setting $B_{m_1 \ldots m_n} = A_{m_1 \ldots m_n}$ (best approximation), and since the l.h.s. of (5.3.12) is non-negative, we also obtain the Bessel inequality:

$$\sum_{m_1, \cdots, m_n = 0}^{n} | A_{m_1 \cdots m_n} |^2 \leqslant \mathbf{E}\{\,| F[W] |^2\}; \quad n = 0, 1, 2, \ldots \tag{5.3.13}$$

5.3c Step 3: Proof of Theorem for special $F(W) = f(V_1, \ldots, V_n)$:

Consider functionals F of the form:

$$F[W] = f(V_1, \ldots, V_n) \tag{5.3.14}$$

where

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_n)\, e^{-(u_1^2 + \cdots + u_n^2)/2}\, du_1 \ldots du_n < \infty . \tag{5.3.15}$$

First, we note that

$$\begin{aligned} A_{m_1 \cdots m_p} &= E\{F[W]\, \Psi_{m_1, \cdots, m_p}(W)\} \\ &= (2\pi)^{-\frac{p}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_n) \left[\prod_{j=1}^{p} H_{m_j}^{N}(u_j) e^{\frac{-u_j^2}{2}} du_j \right], p \geqslant n \end{aligned} \tag{5.3.16}$$

by virtue of (5.3.8). We need to consider two cases: $p = n$ and $p > n$ with $m_p \neq 0$. (Since the m_j's can be zero, we can always arrange to have at least n factors in the product, and so the two cases cover all cases.)

If $p > n$ and $m_p \neq 0$, then the factor

$$\int_{-\infty}^{\infty} H_{m_p}^N \; (u_p) \; e^{\frac{-u_p^2}{2}} \; du_p \tag{5.3.17}$$

comes out, but is zero because of (5.3.10) with $H_0^N \; (u_p) \equiv 1$. Thus,

$$A_{m_1 \cdots m_p} = \begin{cases} 0, & p > n \text{ and } m_p \neq 0, \\ f_{m_1, \cdots, m_n} & p = n \end{cases} \tag{5.3.18}$$

where $f_{m_1, \cdots, m_n}$ is the Hermite coefficient of $f(u_1, \ldots, u_n) \; e^{-(u_1^2 + \cdots + u_n^2)/4}$:

$$f_{m_1, \cdots, m_n} = (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_n) \; e^{-(u_1^2 + \cdots + u_n^2)/4} \times \prod_{j=1}^{n} \left[H_{m_j}^N \, (u_j) \; e^{\frac{-u_j^2}{4}} \; du_j \right]. \tag{5.3.19}$$

Therefore,

$$E \left\{ \left| F[W] - \sum_{m_1, \cdots, m_n = 0}^{N} A_{m_1 \cdots m_n} \Psi_{m_1 \cdots m_n} (W) \right|^2 \right\}$$

$$= (2\pi)^{-\frac{n}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| f(u_1, \; \cdots, u_n) - \sum_{m_1, \cdots, m_n = 0}^{N} f_{m_1 \cdots m_n} \; H_{m_1}^N (u_1) \ldots H_{m_n}^N (u_n) \right|^2 \times e^{-(u_1^2 + \cdots + u_n^2)/2} \; du_1 \ldots du_j \; du_n \, . \tag{5.3.20}$$

Because the Hermite functions are a C.O.N. set of $L_2(-\infty, \infty)$, the r.h.s. of (5.3.20) approaches zero as $N \rightarrow \infty$. But since $A_{m_1 \cdots m_p}$ as defined in (5.3.18) is zero when $p > n$ and $m_p > 0$, we see we can increase the number of subscripts on Ψ and A in (5.3.20) beyond n without changing the sum. We can, therefore, take N subscripts, instead of n, and then (5.3.4) holds for any F defined by (5.3.14). This proves the theorem for the special form of f in (5.3.14).

5.3d Step 4: The special F's are dense in $L_2(W[a, b])$.

As discussed in section 2, P_W is a probability measure on $C[a, b]$, the continuous paths defined on any bounded set $[a, b]$. Denote as $L_2(C)$ all finite mean-square functionals of continuous functions on $[a, b]$ governed by P_W. Any functional of $L_2(C)$ can be approximated by step functionals which, in turn, are finite linear combinations of indicator functionals of W-measurable sets. Because the Wiener process is *separable*, each indicator functional of a W-measurable set can be approximated by a finite linear combination of indicator functionals of windows as described below. [See WIE 2] It therefore only remains to show that the indicator functional of a window can be approximated by the functionals described in step 3.

Let Q denote a *window*. [See Figure 5.3.1]

$$Q = \{W \mid l_j \leqslant W(t_j) \leqslant u_j; \quad j = 1, \ldots, n; \quad a < t_1 < \ldots < t_n \leqslant b\}. \tag{5.3.21}$$

Let $\epsilon > 0$ and let $\rho_{j,\epsilon}(u)$ be a continuous trapezoidal function which is zero outside $(l_j - \epsilon,\ u_j + \epsilon)$, is unity on $[l_j,\ u_j]$, and is linear on the remaining intervals [see Figure 5.3.2].

Let

$$\chi_\epsilon(W) = \prod_{j=1}^{n} \rho_{j,\epsilon}\,[W(t_j)] \tag{5.3.22}$$

and let

$$\chi(W) = \begin{cases} 1, & W \in Q \\ 0 & \text{otherwise} \end{cases} \tag{5.3.23}$$

so that $\chi(W)$ is the indicator of Q. Now,

$$\lim_{\epsilon \to 0} \chi_\epsilon(x) = \chi(x) \tag{5.3.24}$$

for each x in C, and by bounded convergence,

$$\lim_{\epsilon \to 0} \text{q.m. } \chi_\epsilon(W) = \chi(W)\,. \tag{5.3.25}$$

We next want to approximate $W(t_j)$, $j = 1, \ldots n$, using first-order stochastic integrals with $\{\alpha_k(t)\}$ as kernels. Recall $W(a) \equiv 0$. Let

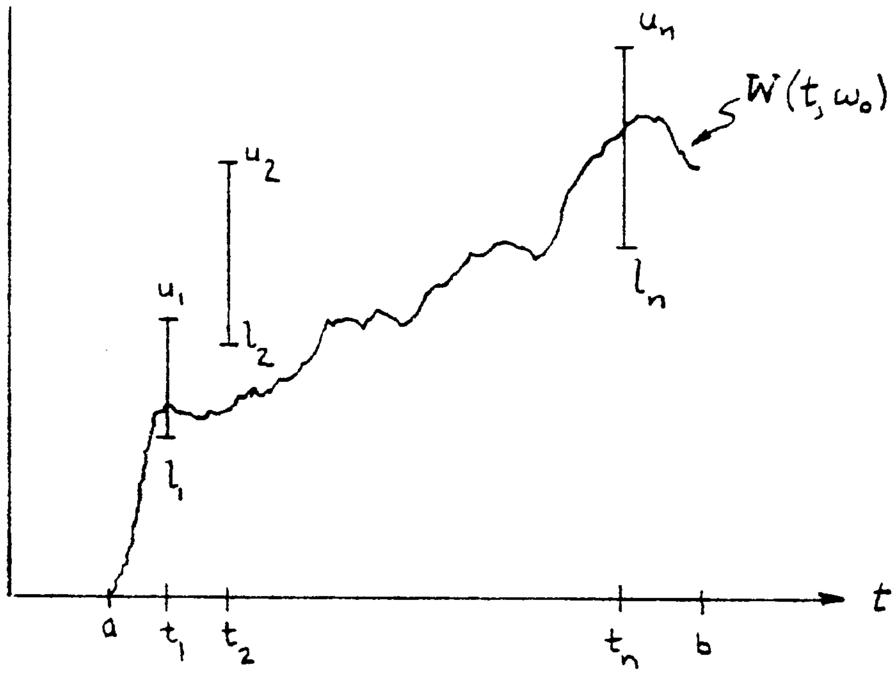


Fig. 5.3.1 — Illustration of a window, Q

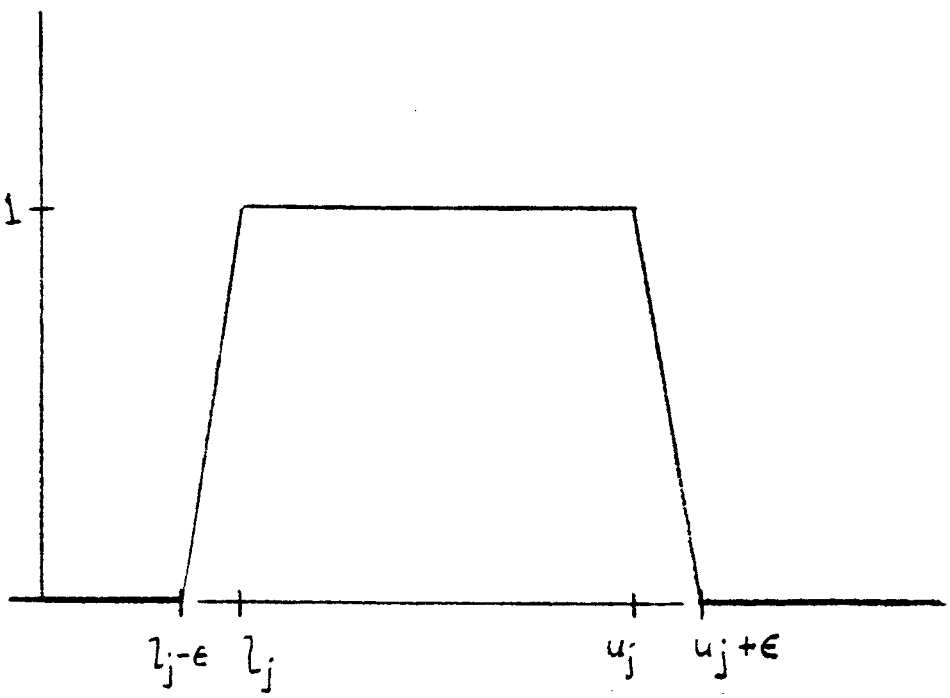


Fig. 5.3.2 — Illustration of trapezoid function, $\rho_{j,\epsilon}$

$$h_j(t) = \begin{cases} 1 & a \leqslant t \leqslant t_j \\ 0 & t_j < t \leqslant b \end{cases} \tag{5.3.26}$$

and let $a_{j,k}$ be the coefficients of the orthogonal development of h_j in terms of $\{\alpha_k(t)\}$:

$$a_{j,k} = \int_a^b h_j(u)\, \alpha_k(u)\, du\,. \tag{5.3.27}$$

Let

$$H_{j,m}(t) = \sum_{k=1}^{m} a_{j,k}\, \alpha_k(t) \tag{5.3.28}$$

so that

$$\underset{m \to \infty}{l.i.m.}\ H_{j,m}(t) = h_j(t) \qquad \text{on } a \leqslant t \leqslant b\,. \tag{5.3.29}$$

Using (2.1.5), (3.3.2) and (3.2.2), we get

$$\mathbf{E}\left[\int_a^b (H_{j,m}(t) - h_j(t))\, dw(t)\right]^2 = \int_a^b [H_{j,m}(t) - h_j(t)]^2\, dt \tag{5.3.30}$$

$$\to 0 \text{ as } m \to \infty\,. \tag{5.3.31}$$

Thus,

$$\int_a^b H_{j,m}(t)\, dw(t) \to \int_a^b h_j(t)\, dw(t) \tag{5.3.32}$$

in the $L_2(C)$ sense on C as $m \to \infty$, and hence there exists a subsequence $m_1, m_2, \ldots$ such that

$$\lim_{p \to \infty} \int_a^b [H_{j,m_p}(t) - h_j(t)]\, dW(t) = 0 \tag{5.3.33}$$

for almost all W in C. We can choose the subsequence such that it holds for all $j = 1, \ldots, n$.

Next, define

$$\chi_{\epsilon,p}(W) = \prod_{j=1}^{n} \rho_{j,\epsilon}\left[\int_a^b H_{j,m_p}(t)\, dW(t)\right]. \tag{5.3.34}$$

For almost all W in C,

$$\lim_{p\to\infty} \chi_{\epsilon,p}(W) = \prod_{j=1}^{n} \rho_{j,\epsilon}\left[\int_a^b h_j(t)\, dW(t)\right]$$
$$= \prod_{j=1}^{n} \rho_{j,\epsilon}\,[W(t_j)]$$
$$= \chi_\epsilon(W)\,. \tag{5.3.35}$$

But since $|\rho_{j,\epsilon}| \leqslant 1$, we have, by bounded convergence,

$$\lim_{p\to\infty} \text{q.m. } \chi_{\epsilon,p}(W) = \chi_\epsilon(W) \text{ on } C\,. \tag{5.3.36}$$

Therefore, $\chi(W)$ can be approximated in the $L_2(C)$ sense by functionals of type $\chi_{\epsilon,p}(W)$. If we choose

$$f(u_1, \cdots u_{m_p}) = \prod_{j=1}^{n} \rho_{j,\epsilon}\left[\sum_{k=1}^{m_p} a_{j,k} u_k\right], \tag{5.3.37}$$

f, which is of the special type in step 3, is therefore shown to be dense in $L_2(C)$.

5.3e Step 5: Completion of Proof

Let $F[W] \in L_2(C)$ and let $A_{m_1 \cdots m_n}$ be its Fourier-Hermite coefficients as in (5.3.5). Given $\epsilon > 0$, and by step 4, there is an $F^*[W]$ satisfying the hypotheses of step 3 such that

$$E\{|F[W] - F^*[W]|^2\} < \frac{\epsilon}{4}\,. \tag{5.3.38}$$

If $A^*_{m_1 \cdots m_n}$ are the Fourier-Hermite coefficients of $F^*[W]$, we can choose N large enough so that

$$E\left\{\left|F^*[W] = \sum_{m_1, \cdots m_N = 0}^{N} A^*_{m_1 \cdots m_N} \Psi_{m_1 \cdots m_N}(W)\right|^2\right\} < \frac{\epsilon}{4}, \tag{5.3.39}$$

so by the Minkowski inequality,

$$E\left\{\left|F[W] - \sum_{m_1, \cdots m_N}^{N} A^*_{m_1 \cdots m_N} \Psi_{m_1 \cdots m_N}(W)\right|^2\right\} < \epsilon\,. \tag{5.3.40}$$

But, the best approximation theorem in Step 2, (5.3.40) must also hold when A^* are replaced by A.

Q.E.D.

5.4 Extensions to the Cameron-Martin Theorem

Point 1: The C-M theorem applies not only to functionals of Wiener processes W, but also to functionals of Wiener increments dW on any closed interval. To see this,

$$W(t) = \int I_{[a,t]} \, d\beta_W \tag{5.4.1}$$

$$= \int_a^t dW(s) \tag{5.4.2}$$

$$= \int_a^t n(s) \, ds , \tag{5.4.3}$$

any Wiener process functional on $[a,b]$ can be expressed as a functional with respect to Wiener measure, Wiener increments, or white noise, respectively, on $[a,b]$. To show the converse, note that the crucial step 4 of the C-M theorem relies on approximating W-window indicators (χ). These can be approximated (or replaced!) by indicators for increments ΔW of the form

$$a_i \leqslant W(t_{i+1}) - W(t_i) \leqslant b_i . \tag{5.4.4}$$

For example, consider the partition of $[l_1, u_1]$ in (5.3.21) as follows:

$$l_1 = x_1 < \ldots < x_{n+1} = u_1 \tag{5.4.5}$$

such that

$$\delta_n = \max_{i=1}^{n} |x_{i+1} - x_i| \rightarrow 0 \text{ as } n \rightarrow \infty . \tag{5.4.6}$$

The set Q corresponding to two windows $[l_1, u_1]$ at t_1, $[l_2, u_2]$ at t_2, can be written as follows:

$$
\begin{aligned}
Q &= \{\omega : l_1 \leqslant W(t_1) \leqslant u_1;\ l_2 \leqslant W(t_2) \leqslant u_2\} \\
&= \bigcap_n \bigcup_{k=1}^{n} \{\omega : x_k \leqslant W(t_1) \leqslant x_{k+1}; \\
& l_2 - x_k \leqslant (W(t_2) - W(t_1)) \leqslant u_2 - x_k\}\,.
\end{aligned} \tag{5.4.7}
$$

This can be extended to any finite partition $\{t_i\}$ of $[a,b]$. By the same argument as in [CAM], it then is only necessary to approximate functionals of the form:

$$
\prod_{i=1}^{n} \chi_{[a_i \leqslant \Delta W(t_i) \leqslant b_i]}, \tag{5.4.8}
$$

where $\Delta W(t_1) = W(t_1)$, $\Delta W(t_2) = W(t_2) - W(t_1)$, etc., by a Fourier-Hermite expansion. This is done by using gate function of intervals $[t_i, t_{i+1})$ instead of h_j in (5.3.26). The rest goes through as before.

The notation $F(dW)$ will be used to emphasize a functional of an increment process in what follows.

Point 2: We may choose $\{\alpha_p(t)\}$ as a C.O.N. set on $L_2(-\infty, \infty)$, rather than $L_2[a, b]$. Recall the purpose of this set is to reconstruct Wiener increments ΔW using first-order stochastic integrals:

$$
\int_a^b \left[\sum_{k=1}^{m} a_{j,k} \alpha_k(t) \right] dW(t) \rightarrow \Delta W(t_j)\,. \tag{5.4.9}
$$

We merely need to approximate gate functions on $[t_i, t_{i+1})$ within the interval $[a, b]$ using a series of the α's, and this can be done using a C.O.N. set defined on $L_2(-\infty, \infty)$.

Point 3: If the functional is *causal*, *causal* α's suffice. Suppose that $a < 0$, $b > 0$, and $F(dW)$ only depends on increments from $a \leqslant t \leqslant 0$. Then, correspondingly, we only need windows on $\Delta W(t_j)$'s for $t_j \leqslant 0$, and the $\Delta W(t_j)$'s are only dependent on $dW(t)$ for $t \leqslant 0$. Therefore, we may then choose

$$
\alpha_k(t) = 0\,, \qquad t > 0\,. \tag{5.4.10}
$$

A suitable choice of C.O.N. set on the half-line would be the Laguerre functions (see section 5.5).

Point 4: To permit $T = (-\infty, \infty)$, the functional F must be restricted. The role of interval $[a, b]$ in the C-M theorem is tied to the ability to approximate the functional F with indicators of finitely many windows of the continuous Wiener paths on a bounded interval. This approximation only holds for functionals on $(-\infty, \infty)$ if the dependence of F on the tails of $W(t)$ (or $dW(t)$) can be bounded.

Let $g_n(t)$ be sequence of widening trapezoid functions:

$$g_n(t) = \begin{cases} 0 & |t| > n + \epsilon \\ 1 & |t| \leqslant n \\ \text{linear} & \text{otherwise} \end{cases} \tag{5.4.11}$$

and suppose

$$F(dW(t)) = \lim_n F(g_n(t)\, dW(t)\,. \tag{5.4.12}$$

Then such a functional admits a Fourier-Hermite expansion. To see this, note that, for some n, F is mean-square approximated by a functional defined on a bounded interval. By choosing windows within this interval, and a C.O.N. set $\{\alpha_p(t)\}$ on $(-\infty, \infty)$ as described in Point 2, the argument of the C-M theorem proceeds as before.

I call functionals of obeying (5.4.12) *lossy* functionals because they lose track of distant portions of $dW(t)$. Any functional defined on an interval $[a, b]$ is automatically lossy.

Point 5: A finite mean-square process $Y(t)$ generated by a stationary lossy L_2 functional F driven by the flow of Wiener increments:

$$Y(t) = F(dW(\cdot + t)) \tag{5.4.13}$$

is *stationary*. This follows directly from the stationarity of the Wiener increments dW. (See 3.4).

Point 6: The adoption of lossy functionals for the sake of modelling processes is not merely desirable for simplifying the representation — it is apparently necessary as well. As McKean points out [MCK1], for an L_2 process Y to admit a Wiener expansion, it must be *mixing*:

$$\lim_{\tau \to \infty} P(A \cap B_{+\tau}) = P(A)\, P(B) \tag{5.4.14}$$

where A and B are events "Y belongs to A or to B" and $B_{+\tau}$ denotes "$Y(\cdot + \tau)$ belongs to B". If the mixing process Y is to be approximated by as a functional F driven by the flow of white noise, F must become independent of remote properties of the white noise, and thus, it must be lossy.

5.5 Laguerre Functions

The Laguerre functions $l_n(x)$ are a C.O.N. set on $(0, \infty)$:

$$\int_0^\infty l_m(x)\, l_n(x)\, dx = \delta_{mn}\, . \tag{5.5.1}$$

They are discussed at considerable length in [SHE1]. They are related to the Laguerre polynominals $L_n(x)$ as follows:

$$l_n(x) = \begin{cases} L_n(x)\, e^{-px} & x \geqslant 0 \\ 0 & x < 0 \end{cases} \tag{5.5.2}$$

$$L_n(x) = \sqrt{2p} \sum_{k=0}^{n} \frac{(-1)^k\, n!}{k!\, [(n-k)\,!]^2} (2px)^{n-k} \tag{5.5.3}$$

The Laguerre functions are perhaps most readily identified via their Laplace transform representation:

$$\Lambda_n(s) = \left[\frac{\sqrt{2p}}{p+s} \right] \left[\frac{p-s}{p+s} \right]^n; \qquad s = \sigma + j\omega,\, \sigma > -p \tag{5.5.4}$$

where it is seen the underlying realization is a simple first-order low pass followed by a cascade of all-pass 180° phase shifters. (See Figure 5.5.1). It is also seen that precisely n integrators are needed to synthesize *all* of the first n orthonormal causal functionals of white noise $n(t)$, that is,

$$\int_{-\infty}^{t} l_i(t-s)\, n(s)\, ds\,, \quad i = 1, 2, \ldots n\,. \tag{5.5.5}$$

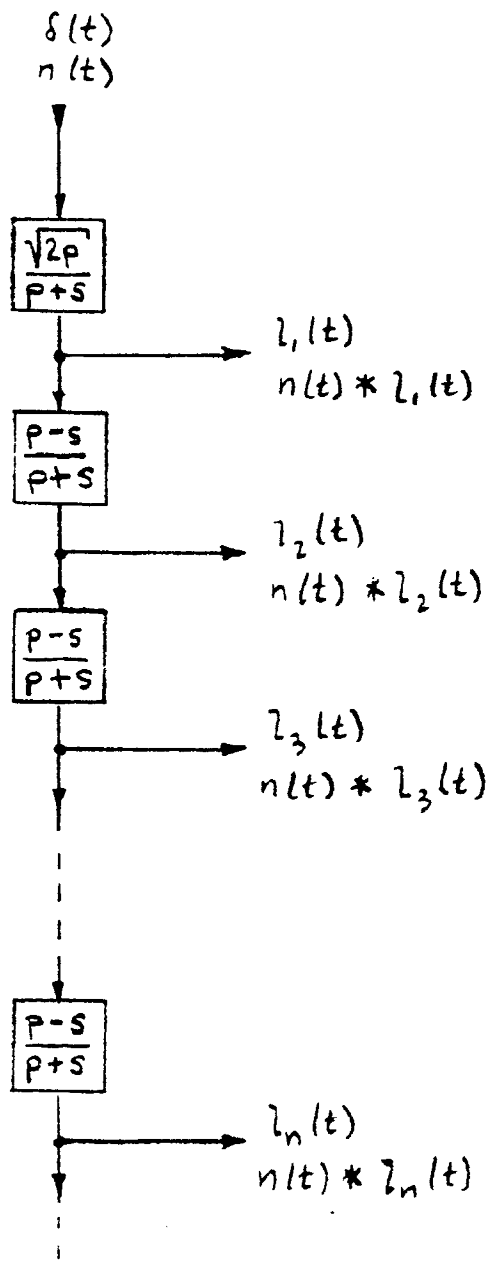


Fig. 5.5.1 — Cascade synthesis of Laguerre linear functionals

"You cannot make a cheap palace." ..... Emerson

"What immortal hand or eye / could frame thy fearful symmetry?" ..... W. Blake

6. Multiple Ito Integrals

The multiple Ito integral, is a crucial part of the process realization theory. It bears an important relationship to both the Fourier-Hermite Expansions and to Volterra Series driven by white noise. A p-th order multiple Ito integral at once realizes the p-th term in Wiener's Homogeneous Chaos [WIE1-2], the expansion of lossy finite mean-square functionals of Wiener processes in terms of orthogonal homogeneous integral operators. The orthogonality is achieved immediately by a construction I call "avoiding diagonals." Although multiple Ito integrals need not be causal functionals, when they are indeed causal, the multiple Ito integrals can be evaluated iteratively as the solution to a nested set of Ito stochastic nonlinear differential equations. The close link to specifically Ito stochastic differential equations is traceable to diagonal avoidance. The p-th order multiple Ito integral is directly equivalent to the p-th order Wiener G functional, which is obtained by orthogonalizing stochastic Volterra series. This connection permits a rigorous discussion of the Lee-Schetzen method for identifying Wiener kernels from input-output cross correlation, and clarifies many mysteries surrounding this identification scheme.

The description of the multiple Ito integral and its properties that follows is based heavily on Ito's own work [ITO1]. The frequent remarks separating the sections tie important properties of these integrals to other topics in this paper.

6.1 Special functions

Let $L_2(T^p)$ denote the collection of square-integrable complex-valued functions on product measure space $(T, B, m)^p$. The function $f(t_1, \ldots, t_p)$ is *special* if

$$f(t_1, \ldots, t_p) = 0 \text{ on diagonals } t_i = t_j \text{ for some } i \neq j. \qquad (6.1.1)$$

The collection of special functions is denoted S_p.

6.2 Theorem: S_p is a linear manifold dense in $L_2(T^p)$.

Proof. We first show the indicator function $\chi_E(t_1, \ldots, t_p)$ of the set $E = E_1 \times \ldots \times E_p$, $E_i \in B^*$ (see section 2) can be approximated in the L_2-norm by a special elementary function. Then, we will have shown it for all L_2 functions on T^p by virtue of the denseness of the indicator functions on sets of finite measure.

By the continuity property of Lebesgue measure m (2.6), (2.7), there exists a collection $\{F_1, \ldots F_n\} \in B^*$ such that

$$F_i \cap F_j = \emptyset, \quad i \neq j, \qquad (6.2.1)$$

$$m(F_i) < \epsilon_1 = \frac{\epsilon}{\binom{p}{2} \left(\sum_i m(E_j)\right)^{p-1}}, \qquad (6.2.2)$$

and

$$E_i = \bigcup_j F_{u_j}, \quad i = 1, \ldots, p. \qquad (6.2.3)$$

Thus,

$$\chi_E(t_1, \ldots, t_p) = \sum \epsilon_{i_1 \ldots i_p} \chi_{F_{i_1}}(t_1 \ldots \chi_{F_{i_p}}(t_p), \qquad (6.2.4)$$

where

$$\epsilon_{i_1} \ldots {i_p} = 0 \text{ or } 1. \qquad (6.2.5)$$

We now separate the sum (6.2.4) into two parts

$\sum'$: where the indices $\{i_1, \ldots, i_p\}$ are all different, and (6.2.6)
$\sum''$: the remainder.

Let

$$f(t_1, \ldots, t_p) = \sum' \epsilon_{i_1, \ldots, i_p} \chi_{F_{i_1}}(t_1) \ldots \chi_{F_{i_p}}(t_p). \qquad (6.2.7)$$

Then $f \in S_p$ and

$$||\chi_E - f||^2 = \int \ldots \int \left|\chi_E(t_1, \ldots, t_p) - f(t_1, \ldots, t_p)^2 m(dt_1) \ldots m(dt_p)\right.$$

$$= \sum{}'' \epsilon_{i_1, \ldots, i_p} m(F_{i_1}) \ldots m(F_{i_p}) \tag{6.2.8}$$

Now, $\epsilon_{i_1, \ldots, i_p} = 0$ or 1, and $\sum''$ contains terms with pairs of identical indices. Each pair (F_i, F_i) occurs (at least) $\binom{p}{2}$ times. Thus,

$$|| \chi_E - f||^2 \leqslant \binom{p}{2} \sum_i (m(F_i))^2 \sum_{j_1, \cdots j_{p-2}} m(F_{j_1}) \ldots m(F_{j_p-2}) \tag{6.2.9}$$

where the last sum in (6.2.9) is simply equal to

$$\left(\sum_j m(f_j)\right)^{p-2}. \tag{6.2.10}$$

Moreover, $m(F_i) < \epsilon_1$, so

$$||\chi_E - f||^2 \leqslant \binom{p}{2} \sum_i \epsilon_1 m(F_i) \left(\sum_j m(F_j)\right)^{p-2} \tag{6.2.11}$$

$$= \binom{p}{2} \epsilon_1 \left(\sum_j m(F_j)\right)^{p-1} \tag{6.2.12}$$

$$= \epsilon . \tag{6.2.13}$$

Q.E.D.

Figure (6.2.1.) illustrates this construction for the case of $p = 2$, using intervals for the sets F_i. The blackened rectangles represent pairs $F_i \times F_j$ belonging to the sum $\sum''$. Rectangles $F_i \times F_j$ corresponding to $\epsilon_{i_1, i_2} = 0$ are not shown. (They are outside E.) Note how the special functions avoid the diagonal $t_1 = t_2$.

While special functions are indeed dense in $L^2(T^p)$, this does not mean that the sum $\sum''$ over the diagonals, the discarded sum, has little value or consequence! The discarded $\sum''$ is precisely what distinguishes the multiple Ito integral from a Volterra integral driven by white noise, and what gives rise to the multiple Ito integral's orthogonality property. In essence, by throwing $\sum''$ away, the construction of the multiple

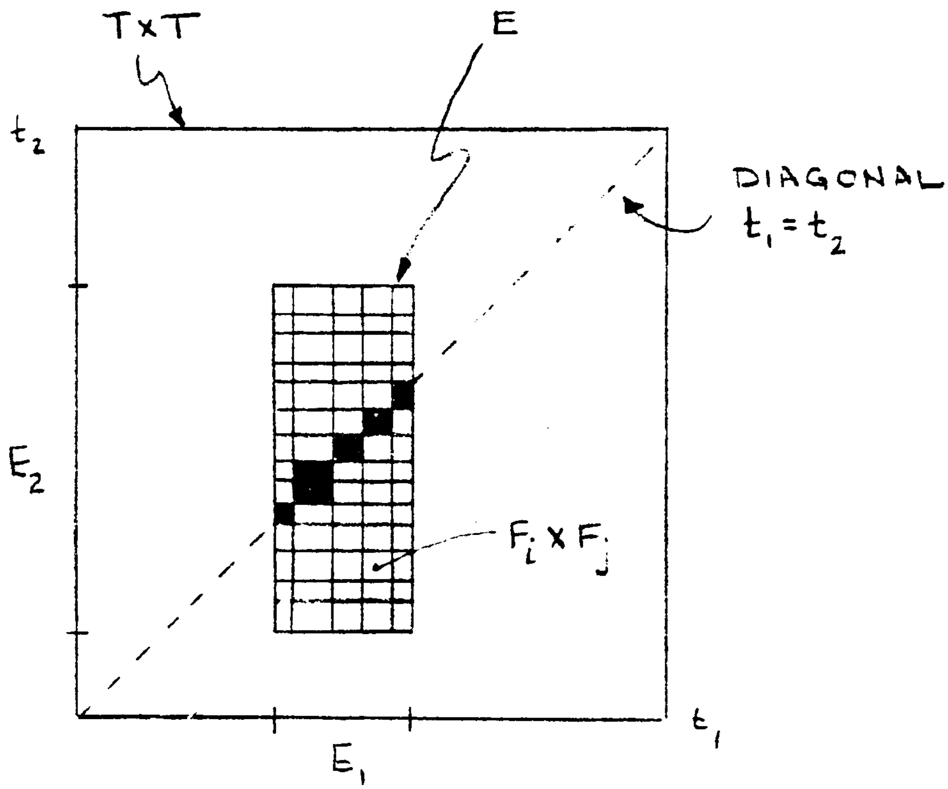


Fig. 6.2.1 — Illustration of special set construction, $p = 2$

Ito integral bypasses exactly those p-th order product terms that collapse onto lower order term subspaces by virtue of the property

$$dW_t^2 \approx dt\,, \tag{6.2.14}$$

which manifests itself only along diagonals.

Since the special functions avoid diagonals entirely, the multiple Ito integral does not depend on values of its kernel f along diagonals. The values of the kernel f along diagonals may be assigned arbitrarily without affecting the integral or its statistical properties. This, in turn, means that no scheme can identify the kernel along diagonals from statistical or input-output properties of the multiple Ito integral. Furthermore, for the purpose of constructing process representations, it is not necessary to identify the diagonal values of the kernel at all!

6.3 Multiple Ito Integrals, f = Special Elementary Function

Let

$$f(t_1, \ldots, t_p) = \begin{cases} a_{i_1}, & (t_1, \ldots t_p) \in T_{i_1} \times \ldots \times T_{i_p} \\ 0 & \text{elsewhere,} \end{cases} \tag{6.3.1}$$

where $T_1, \ldots, T_n$ are disjoint, $m(T_i) < \infty$, $i = 1, \ldots, n$, and $a_{i_1} \ldots i_p = 0$ if any two of $i_1, \ldots i_p$ are equal. Then

$$\begin{aligned} I_p(f) &= \int \ldots \int f(t_1, \ldots, t_p)\, dW(t_1) \ldots dW(t_p) \\ &\triangleq \sum_{i_1, \ldots, i_p} a_{i_1, \ldots, i_p}\, \beta(T_{i_1}) \ldots \beta(T_{i_p})\,, \end{aligned} \tag{6.3.2}$$

where B is the Wiener random measure associated with measure m = Lebesque measure.

6.4 Theorem: Linearity

$$I_p(af + bg) = aI_p(f) + bI_p(g)\,, \quad f, g \text{ elementary.} \tag{6.4.1}$$

6.5 Theorem: Symmetrization

$$\text{Let } \tilde{f}(t_1, \ldots, t_p) = \frac{1}{p!} \sum_{(\pi)} f(t_{\pi_1}, \ldots, t_{\pi_p}) \tag{6.5.1}$$

where $(\pi) = (\pi_1, \ldots, \pi_p)$ running over all permutations of $(1, 2, \ldots, p)$. That is, $\tilde{f}$ is the symmetrization of f. Then

$$I_p(f) = I_p(\tilde{f}). \tag{6.5.2}$$

Proof. Let $(j) \sim (i)$ mean $(j_1, \ldots, j_p)$ is a permutation of $(i_1, \ldots, i_p)$. Now, in the sum (6.3.2) we identify all terms with the identical product $\beta(T_{i_1}) \ldots \beta(T_{i_p})$. Since the order of these factors does not matter, we see that these products occur along with $a_{j_1}, \ldots, j_p$ in every possible $(j) \sim (i)$. Thus, by re-grouping terms,

$$I_p(f) = \sum_{i_1 < i_2 < \cdots i_p} \left[\sum_{(j) \sim (i)} a_{j_1, \cdots j_p} \right] \beta(T_{i_1}) \ldots \beta(T_{i_p}) \tag{6.5.3}$$

$$= p! \sum_{i_1 < i_2 < \cdots i_p} \left[\frac{1}{p!} \sum_{(j) \sim (i)} Q_{j_1 \cdots j_p} \right] \beta(T_{i_1}) \ldots \beta(T_{i_p}). \tag{6.5.4}$$

The inner term, viewed as a function of the i's, does not depend on the specific order of the i's in the outer sum. Thus, we may now re-introduce an un-ordered outer sum, being mindful that $a_{i_1, \ldots, i_p} = 0$ if any two i's are equal, and that an un-ordered sum itself gives rise to $p!$ occurrences of each combination of the i's:

$$I_p(f) = \sum_{i_1, \cdots, i_p} \left[\frac{1}{p!} \sum_{(j) \sim (i)} a_{j_1, \cdots, j_p} \right] \beta(T_{i_1}) \ldots \beta(T_{i_p})$$

$$= I_p(\tilde{f}). \tag{6.5.5}$$

Q.E.D.

This theorem and theorem 6.13 show that the Ito kernels can only be distinguished from the statistics of the Ito integral modulo the symmetrization. This leads to the important conclusion that the Lee-Schetzen identifies the symmetric form of the kernel. In system realization, this creates a frustration in two ways. The first is that it makes it difficult to recognize possibly simpler, though non-symmetric, forms of the kernel which correspond to lower order systems. The second is that symmetrization is a neccessary step to realizing stationary, causal Ito integrals by stationary bilinear systems, a step that greatly enlarges the required state space dimension. This is discussed further in section 8.

In the next three theorems, multiple Ito integrals of varying order are shown to be orthogonal to one another, and an isomorphic relationship with symmetric functions $\tilde{f} \in L_2(T^p)$ is established. This is a principal benefit of the multiple Ito integral. In contrast, multiple Volterra functionals with $L_2(T^p)$ kernels driven by white noise are not orthogonal and do not adhere to the isomorphism as in Theorem 6.8. In Section 7, it will be shown that the multiple Ito integrals of p-th order are equivalent to a p-th order Volterra functional *series* obtained by Gram-Schmidt orthogonalization of the leading p-th order Volterra functional operator.

6.6 Theorem: Inner Product of Multiple Ito Integrals of Equal Order

$$\mathrm{E}\{I_p(f)\, I_p{}^*(g)\} = p! \int \dots \int \tilde{f}(t_1, \dots, t_p)\, \tilde{g}^*(t_1, \dots, t_p)\, m(dT_1) \dots m(dT_p) \,. \tag{6.6.1}$$

Proof. Recall that

$$\mathrm{E}\{\beta(T_i)\, \beta(T_j)\} = \begin{cases} m(T_i) & i = j \\ 0 & i \neq j \end{cases} . \tag{6.6.2}$$

Therefore,

$$\mathrm{E}\{I_p(f)\; I^*_p(g)\} = \mathrm{E}\left\{ \sum_{i_1 < \dots < i_p} \left(\sum_{(j) \sim (i)} a_{j_1 \dots j_p} \right) \beta(T_{i_1}) \dots \beta(T_{i_p}) \right] \cdot$$
$$\left[\sum_{k_1 < \dots < k_p} \left(\sum_{(l) \sim (k)} b_{l_1 \dots l_p} \right) \beta(T_{k_1}) \dots \beta(T_{k_p}) \right]^* \Bigg\} \,. \tag{6.6.3}$$

We can always refine the sets T_i so that the same sets apply to both f and g. Then by (6.6.2),

$$\mathrm{E}\{I_p(f)\, I^*_p(g)\} = \sum_{i_1 < \ldots < i_p} \left(\sum_{(j) \sim (i)} a_{j_1 \ldots j_p} \right)$$

$$\left(\sum_{(k) \sim (i)} b^*_{k_1 \ldots k_p} \right) m(T_{i_1}) \ldots m(T_{i_p}), \tag{6.6.4}$$

that is, all cross terms vanish. Using the same trick as in (6.5), we now replace the ordered outer sum with a un-ordered sum:

$$\mathrm{E}\{I_p(f)\, I_p{}^*(g)\} = \frac{1}{p!} \sum_{i_1, \ldots, i_p} (\cdots)(\cdots)\, m(T_{i_1}) \ldots m(T_{i_p}) \tag{6.6.5}$$

$$= p! \sum_{i_1, \ldots i_p} \left[\frac{1}{p!} \cdots \right] \left[\frac{1}{p!} \cdots \right] m(T_{i_1}) \ldots m(T_{i_p}) \tag{6.6.6}$$

$$= p! \int \ldots \int \tilde{f}(t_1, \ldots, t_p) \cdot$$

$$\tilde{g}^*(t_1, \ldots, t_p)\, m(dt_1) \ldots - m(dt_p) \tag{6.6.7}$$

Q.E.D.

6.7 Theorem: Inner product of Multiple Ito Integrals of Different Order

$$\mathrm{E}\{I_p(f)\, I^*_q(g)\} = 0, \qquad p \neq q. \tag{6.7.1}$$

Proof: Assume $p > q$. Then, in writing the formula in an expansion of the form (6.6.3), each term is seen to contain $\beta(T_i)$ factors from I_p that are always unpaired with other factors $\beta(T_j)$ and $\beta(T_k)$ from both I_p and I_q respectively. The conclusion then follows immediately from the fact that distinct β's are independent, and that $\mathrm{E}\{\beta(T_i)\} = 0$.

Q.E.D.

6.8 Theorem: Variance of Multiple Ito Integral

$$\mathrm{E}\{|I_p(f)|^2\} = p! \int \ldots \int |\tilde{f}(t_1, \ldots, t_p)|^2\, m(dt_1) \ldots m(dt_p) \tag{6.8.1}$$

$$\leqslant p! \int \ldots \int |f(t_1, \ldots, t_p|^2\, m(dt_1) \ldots m(dt_p). \tag{6.8.2}$$

Proof: (6.8.1) follows immediately from (6.6.1) with $g = f$, while (6.8.2) holds by virtue of Schwarz's inequality.

Q.E.D.

6.9 Definition: Multiple Ito Integral, $f \in L_2(T^p)$

Theorems 6.4 through 6.8 show that I_p may be regarded as a bounded linear operator from S_p into $L_2(\omega)$. By extension via a sequence of integrals of special elementary functions converging to f,

$$I_p(f) \triangleq \lim_{n\to\infty} \text{q.m. } I_p(f_n) \tag{6.9.1}$$

where $||f - f_n|| \to 0$, it becomes an operator from the closure of S_p, which by Theorem 6.2 is $L_2(T^p)$, into $L_2(\omega)$ also satisfying Theorems 6.4 through 6.8.

By letting L_0^2 denote the complex members and $I_0(c) \triangleq c$, then Theorems 6.4 through 6.8 also hold for $p, q = 0$ as well as $1, 2, \ldots$.

6.10 Recursion Relation

This important theorem is the basis for proving Theorem 6.11 (connection between elementary multiple Ito integrals and Fourier-Hermite functionals). It clarifies how, in the mean square sense, $dW_t^2 \approx dt$ along diagonals, and lays a foundation to show how the avoidance (or masking) of the diagonals inherent in the multiple Ito integral can be represented by a telescoping series of Volterra multiple integrals driven by white noise. This will be the topic of section 7.

6.10a Lemma

If $\rho(t_1, \ldots, t_p) \in L_2(T^p)$ and $\phi(t) \in L_2(T)$, then

$$\int \cdots \left[\int |\rho(t_1, \ldots, t_p)\, \phi(t_k)|\, m(dt_k)\right]^2 m(dt_1) \ldots m(dt_{k-1})\, m(dt_{k+1}) \ldots m(dt_p) \leqslant ||\phi||^2 \cdot ||\rho||^2 < \infty . \tag{6.10.1}$$

Proof:

$$\left[\int |\rho(t_1, \ldots, t_k, \ldots, t_p)\, \phi(t_k)|\, m(dt_k)\right]^2 \leqslant \int |\phi(t_k)|^2\, m(dt_k) \int |\rho(t_1, \ldots, t_k, \ldots, t_p)|^2\, m(dt_k) \text{ a.e.} \tag{6.10.2}$$

by virtue of the Canchy-Schwartz inequality applied to ϕ and to ρ viewed as a function of t_k only.

6.10b. Lemma

The function

$$\rho \underset{(k)}{\times} \phi(t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_p) \equiv \int \rho(t_1, \ldots, t_k, \ldots, t_p)\, \phi\,(t_k)\; m(dt_k) \quad (6.10.3)$$

is a square-integrable function of $t_1, \ldots, t_{k-1}, t_{k+1}, \ldots, t_p$, and

$$||\rho \underset{(k)}{\times} \phi|| \leqslant ||\rho|| \cdot ||\phi|| . \quad (6.10.4)$$

Proof:

Follows immediately from 6.10a, since

$$||\rho \underset{(k)}{\times} \phi||^2 = \int \ldots [\int \rho(t_1, \ldots, t_p)]\; \phi(t_k)\; m(dt_k)]^2 \cdot m(dt_1) \ldots m(dt_{k-1})\; m(dt_{k+1}) \ldots m(dt_p) . \quad (6.10.5)$$

Q.E.D.

6.10c Theorem: Recursion Relation

If $\rho(t_1, \ldots, t_p) \in L_2(T^p)$ and $\phi(t) \in L_2(T)$, then

$$I_{p+1}(\rho\phi) = I_p(\rho) I_1(\rho) - \sum_{k=1}^{p} \cdot I_{p-1}(\rho \underset{(k)}{\times} \phi) . \quad (6.10.6)$$

Proof:

First, let ρ and ϕ be special elementary functions:

$$\rho(t_1, \ldots, t_p) = \begin{cases} a_{i_1 \ldots i_p}, & (t_1, \ldots, t_p) \in T_{i_1} \times \ldots \times T_{ip} \\ 0 & \text{otherwise} \end{cases} , \quad (6.10.7)$$

$$\phi(t) = \begin{cases} b_i, & t \in T_i \\ 0 & \text{otherwise} \end{cases}, \tag{6.10.8}$$

where $T_1, \ldots, T_N$ are disjoint, $M(T_i) < \infty$; $i = 1, \ldots, N$, and $a_{i_1 \ldots i_p} = 0$ if any two of $i_1, \ldots, i_p$ are equal.

Let $S = T_1 + \ldots + T_N$, $A = \max_i |a_i|$, $B = \max_i |b_i|$. Then $m(S) < \infty$, $A < \infty$, $B < \infty$. In addition, by continuity on Lebesgue measure m, we may subdivide the T_i's if necessary so that, for any $\epsilon > 0$,

$$m(T_i) < \epsilon, \quad i = 1, \ldots, N. \tag{6.10.9}$$

Note that S, A and B do not change under this subdivision.

We now introduce a new special elementary function:

$$\chi_\epsilon(t_1, \ldots, t_p, t) = \begin{cases} a_{i_1 \ldots i_p} b_i, & (t_1, \ldots, t_p, t) \in T_{i_1} \times \ldots \times T_{i_p} \times T_i,\ i \neq i_1, \ldots, ip \\ 0 & \text{otherwise} \end{cases} \tag{6.10.10}$$

Then, we have

$$I_p(\rho) \cdot I_1(\phi) = \sum_{i_1, \ldots, i_p} a_{i_1 \ldots i_p} \beta(T_{i_1}) \ldots \beta(T_{i_p}) \sum_i b_i \beta(T_i) \tag{6.10.11}$$

$$= \sum_{i \neq i_1, \ldots, i_p} a_{i_1 \ldots i_p} b_i \beta(T_{i_1}) \ldots \beta(T_{i_p}) \beta(T_i)$$

$$+ \sum_{k=1}^{p} \sum_{i_1, \ldots i_p} a_{i_1 \ldots i_p} b_{i_k} \beta(T_{i_1}) \ldots \beta(T_{i_{k-1}}) \beta(T_{i_k})^2 \beta(T_{i_{k+1}}) \ldots \beta(T_{i_p}) \tag{6.10.12}$$

$$= I_{p+1}(\chi_\epsilon) + \sum_{k=1}^{p} \sum_{i_1, \ldots, i_p} a_{i_1 \ldots i_p} b_{i_k} \beta(T_{i_1}) \ldots \beta(T_{i_{k-1}}) m(T_{i_k}) \beta(T_{i_{k+1}}) \ldots \beta(T_{ip})$$

$$+ \sum_{k=1}^{p} \left[\sum_{i_1 \ldots i_p} a_{i_1 \ldots i_p} b_{i_k} \beta(T_{i_1}) \ldots \beta(T_{i_{k-1}}) \left(\beta(T_{i_k})^2 - m(T_{i_k})\right) \cdot \beta(T_{i_{k+1}}) \ldots \beta(T_{i_p}) \right] \tag{6.10.13}$$

$$= I_{p+1}(\chi_\epsilon) + \sum_{k=1}^{p} I_{p-1}(\rho \underset{(k)}{\times} \phi) + \sum_k R_k, \tag{6.10.14}$$

where R_k denotes the term inside brackets in (6.10.13).

We now show that, as $\epsilon \to 0$,

$$I_{p+1}(\chi_\epsilon) \xrightarrow{\text{m.s.}} I_{p+1}(\rho\phi) \tag{6.10.15}$$

and

$$R_k \xrightarrow{\text{m.s.}} 0\,. \tag{6.10.16}$$

By theorem 6.6,

$$||I_{p+1}(\chi_\epsilon) - I_{p+1}(\rho\phi)||^2 = p!\ ||\chi_\epsilon - \rho \cdot \phi||^2. \tag{6.10.17}$$

Now,

$$\chi_\epsilon - \rho \cdot \phi = \begin{cases} -a_{i_1 \ldots i_k \ldots i_p}\, b_{i_k}, & (t_1, \ldots, t_k, \ldots, t_p, t) \in T_{i_1} \times \ldots \times T_{i_k} \times \ldots \times T_{i_p} \times T_{i_k} \\ 0 & \text{otherwise}, \end{cases} \tag{6.10.18}$$

so

$$p!\ ||\chi_\epsilon - \rho \cdot \phi||^2 \leqslant p! \sum_{k=1}^{p} \sum_{i_1, \ldots, i_p} a_{i_1 \ldots i_p}^2\, b_{i_k}^2 \cdot$$
$$m(T_{i_1}) \ldots m(T_{i_{k-1}})\ m(T_{i_k})^2 \ldots m(T_{i_p}) \tag{6.10.19}$$

$$\leqslant p!\ A^2B^2 \sum_{k=1}^{p} m(T_{i_1}) \ldots m(T_{i_{k-1}})\ m(T_{i_k})^2 \ldots m(T_{i_p}) \tag{6.10.20}$$

$$= p!\ A^2B^2 \sum_{k=1}^{p} \sum_{i_k} m(T_{i_k})^2 \cdot$$
$$\sum_{i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_p} m(T_{i_1}) \ldots m(T_{i_{k-1}})\ m(T_{i_{k+1}})\ m(T_{i_p}) \tag{6.10.21}$$

$$\leqslant p!\ A^2B^2 \left[\sum_i m(T_i)\right]^{p-1} p \left[\sum_i m(T_i)\right]^2 \tag{6.10.22}$$

$$\leqslant p!\ A^2B^2 \left[\sum_i m(T_i)\right]^p p\epsilon \tag{6.10.23}$$

$$\leqslant pp!\ A^2B^2\ m(S)^p\, \epsilon\,. \tag{6.10.24}$$

Since reducing ϵ (by partitioning the T's into smaller pieces) does not affect $m(S)$, A, or B, then

$$||\chi_\epsilon\ \rho\phi||^2 \to 0 \tag{6.10.25}$$

and therefore

$$||I_{p+1}(\chi_\epsilon) - I_{p+1}(\rho\phi)||^2 \to 0\,. \tag{6.10.26}$$

Now, to bound $||R_k||^2$, we note that:

- $\beta(T_k)$ is independent of $\beta(T_j)$, $j \neq k$.
- $\beta^2(T_k) - m(T_k)$ is zero mean and also independent of $\beta(T_j)$, $j \neq k$.
- $\beta(T_k)$ is Gaussian, zero mean, variance $m(T_k)$.
- $$\mathbf{E}\left\{[\beta^2(T_k) - m(T_k)]^2\right\} = m(T_k)^2\, \mathbf{E}\left\{\left[\frac{\beta(T_k)}{\sqrt{m(T_k)}} - 1\right]^2\right\}. \tag{6.10.27}$$
- $\beta(T_k)/\sqrt{m(T_k)}$ is normal Gaussian, zero mean.
- $$\mathbf{E}\left\{[\beta^2(T_k) - m(T_k)]^2\right\} = m(T_k)^2\, \frac{1}{2\pi} \int_{-\infty}^{\infty} (x^2 - 1)^2\, e^{\frac{-x^2}{2}}\, dx \tag{6.10.28}$$

$$= m(T_k)^2 C < \infty. \tag{6.10.29}$$

Thus,

$$||R_k||^2 = C \sum a^2_{i_1 \ldots i_p}\, b^2_{i_k}\, m(T_{i_1}) \cdots m(T_{i_{k-1}})\, m(T_{i_k})^2 \ldots m(T_{ip}) \tag{6.10.30}$$

$$\leqslant CA^2B^2 \epsilon\, m(S)^p. \tag{6.10.31}$$

By making finer partitions of the T_i's, $\epsilon \rightarrow 0$, which means

$$||R_k||^2 \rightarrow 0. \tag{6.10.32}$$

This proves the theorem for special elementary functions.

Next, let ρ and ϕ be any functions in $L^2(T^p)$ and $L^2(T)$ respectively. By theorem 6.2, we can find sequences of special elementary functions $\rho_n \in L^2(T^p)$ and $\phi_n \in L^2(T)$ such that

$$||\rho_n - \rho|| \rightarrow 0 \tag{6.10.33}$$

and

$$||\phi_n - \phi|| \rightarrow 0. \tag{6.10.34}$$

By the above, we have

$$I_{p+1}(\rho_n \phi_n) = I_p(\rho_n)\ I_1(\phi_n) - \sum_{k=1}^{p} I_{p-1}(\rho_n \underset{(k)}{\times} \phi_n)\,. \tag{6.10.35}$$

By using (6.10.1), (6.10.4), and (6.8.2), we get:

$$||I_{p+1}(\rho_n \phi_n) - I_p(\rho\phi)||_{L_1} = ||I_{p+1}(\rho_n \phi_n - \rho\phi)||_{L_1} \tag{6.10.36}$$

$$\leqslant ||I_{p+1}(\rho_n \phi_n - \rho\phi)|| \tag{6.10.37}$$

$$\leqslant \sqrt{(p+1)!}\ ||\rho_n \phi_n - \rho\phi|| \tag{6.10.38}$$

$$\leqslant \sqrt{(p+1)!}\ ||\rho_n(\phi_n - \phi)|| + \sqrt{(p+1)!}\ ||(\rho_n - \rho)\ \phi\ || \tag{6.10.39}$$

$$= \sqrt{(p+1)!}\ ||\rho_n|| \cdot ||\phi_n - \phi|| + \sqrt{(p+1)!}\ ||\rho_n - \rho|| \cdot ||\phi|| \tag{6.10.40}$$

$$\to 0 \text{ as } n \to \infty\,.$$

Also,

$$||I_p(\rho_n)\ I_1(\phi_n) - I_p(\rho)\ I_1(\phi)||_{L_1} \leqslant ||I_p(\rho_n)\ I_1(\phi_n - \phi)||_{L_1}$$
$$+ ||I_p(\rho_n - \rho)\ I_1(\phi)||_{L_1} \tag{6.10.41}$$

$$\leqslant ||I_p(\rho_n)|| \cdot ||I_1(\phi_n - \phi)|| + ||I_p(\rho_n - \rho)|| \cdot ||I_1(\phi)|| \tag{6.10.42}$$

$$\leqslant \sqrt{p!}\ ||\rho_n|| \cdot ||\phi_n - \phi|| + \sqrt{p!}\ ||\rho_n - \rho|| \cdot ||\phi|| \tag{6.10.43}$$

$$\to 0 \text{ as } n \to \infty\,. \tag{6.10.44}$$

And finally,

$$||I_{p-1}(\rho_n \underset{(k)}{\times} \phi_n) - I_{p-1}(\rho \underset{(k)}{\times} \phi)||_{L_1}$$

$$\leqslant ||I_{p-1}(\rho_n \underset{(k)}{\times} \phi_n - \rho \underset{(k)}{\times} \phi)|| \tag{6.10.45}$$

$$\leqslant \sqrt{(p-1)!}\ ||\rho_n \underset{(k)}{\times} \phi_n - \rho \underset{(k)}{\times} \phi|| \tag{6.10.46}$$

$$\leqslant \sqrt{(p-1)!}\ ||(\rho_n - \rho) \underset{(k)}{\times} \phi_n|| + \sqrt{(p-1)!}\ ||\ \rho \underset{(k)}{\times} (\phi_n - \phi)|| \tag{6.10.47}$$

$$\leqslant \sqrt{(p-1)!}\ ||\rho_n - \rho|| \cdot ||\phi_n|| + \sqrt{(p-1)!}\ ||\rho|| \cdot ||\phi_n - \phi|| \tag{6.10.48}$$

$$\to 0 \text{ as } n \to \infty\,. \tag{6.10.49}$$

Q.E.D.

6.11 Theorem: Relationship Between Multiple Ito Integrals and Fourier-Hermite Expansions

Let $\rho_1(t), \rho_2(t), \ldots \rho_n(t)$ be an orthonormal set of real-valued functions in $L_2(T)$ and let $H_p(\chi)$ be the type of p-th degree Hermite polynomial defined in table 5.2.1b. Then,

$$\int \cdots \int \overbrace{\rho_1(t_1)\ldots\rho_1(t_{p_1})}^{p_1 \text{ terms}} \; \overbrace{\rho_2(t_{p_1+1})\ldots\rho_2(t_{p_1+p_2})}^{p_2 \text{ terms}} \ldots \overbrace{\rho_n(t_{p_1+\ldots+p_{n-1}+1})\ldots\rho_n(t_{p_1+\ldots+p_n})}^{p_n \text{ terms}}$$
$$dW(t_1)\ldots dW(t_{p_1+\ldots+p_n})$$
$$= \prod_{k=1}^{n} \frac{H_{p_k}\left[\frac{1}{\sqrt{2}} \int \rho_k(t)\, dW(t)\right]}{(\sqrt{2})^{p_k}} . \tag{6.11.1}$$

Proof:

(By induction with respect to the sum $p_1+\ldots+p_n$). By convention, let $I_0(1) \equiv 1$. The proof is true trivially for both $p_1+\ldots+p_n = 0$ and $p_1+\ldots+p_n = 1$.

Let us now assume that (6.11.1) is true for $p_1+\ldots+p_n = p-1,\, p$. Without loss of generality, let $p_1 \geqslant 1$. In preparation of using theorem 6.10, let

$$\rho(t_1,\ldots,t_p) = \overbrace{\rho_1(t_1)\ldots\rho_1(t_{p_1-1})}^{p_1-1 \text{ terms}} \; \overbrace{\rho_2(t_{p_1})\ldots\rho_2(t_{p_1+p_2-1})}^{p_2 \text{ terms}}$$
$$\ldots \overbrace{\rho_n(t_{p_1}+\ldots+p_{n-1})\ldots\rho_n(t_{p_1}+\ldots+p_n-1)}^{p_n \text{ terms}} , \tag{6.11.2}$$
$$\phi(t) = \rho_1(t) . \tag{6.11.3}$$

From theorem 6.10,

$$\int \cdots \int \rho(t_1,\ldots,t_p)\, \rho_1(t)\, dW(t_1)\ldots dW(t_p)\, dW(t)$$
$$= \int \cdots \int \rho(t_1,\ldots,t_p)\, dW(t_1)\ldots dW(t_p) \int \rho_1(t)\, dW(t)$$
$$- \sum_{k=1}^{p} \int \cdots \int \rho \underset{(k)}{\times} \rho_1\, dW(t_1)\ldots dW(t_{k-1})\, dW(t_{k+1})\, dW(t_p) . \tag{6.11.4}$$

By the orthonormality of $\{\rho_1, \ldots, \rho_N\}$,

$$\rho \underset{(k)}{\times} \rho_1 = \int \rho(t_1,\ldots,t_p)\, \rho_1(t_k)\; m(dt_k)$$

$$= \begin{cases} \rho_1(t_1)\ldots\rho_1(t_{k-1})\,\rho_1(t_{k+1})\ldots\rho_1(t_{p_1-1})\cdot \\ \rho_2(t_{p_1})\ldots\rho_n(t_{p_1+\ldots+p_n-1}), & 1 \leqslant k \leqslant p_1-1, \\ 0 & k \geqslant p_1. \end{cases}$$

We may now evaluate the r.h.s. terms in (6.11.4), using the assumption of (6.11.1) for $p, p-1$ (instead of $p+1$, which we seek to prove). Note order $p_1 + \ldots + p_n - 1 = p$, and $p_1 - 2 + p_2 + \ldots p_n = p - 1$.

$$\int\cdots\int \rho(t_1,\ldots,t_p)\; dw(t_1)\ldots dw(t_p)$$

$$= \int\cdots\int \overbrace{\rho_1(t_1)\ldots\rho_1(t_{p_1-1})}^{p_1-1 \text{ terms}}\;\overbrace{\rho_2(t_{p_1})\;\ldots\;\rho_2(t_{p_1+p_2-1})}^{p_2 \text{ terms}}\cdots$$

$$\cdots\overbrace{\rho_n(t_{p_1+\ldots+p_{n-1}})\cdots\rho_n(t_{p_1+\ldots+p_n-1})}^{p_n \text{ terms}}\; dW(t_1)\ldots dW(t_p)$$

$$= \left[\prod_{i=2}^{n} \frac{H_{p_k}\left[\frac{1}{\sqrt{2}}\int \rho_i\, dW(t)\right]}{(\sqrt{2})^{p_i}}\right] \frac{H_{(p_1-1)}\left[\frac{1}{\sqrt{2}}\int \rho_1(t)\, dW(t)\right]}{(\sqrt{2})^{p_1-1}}. \tag{6.11.6}$$

$$\int\cdots\int \rho \underset{(k)}{\times} \phi_1\; dW(t_1)\ldots dW(t_{k-1})\, dW(t_{k+1})\ldots dW(t_p) \tag{6.11.7}$$

$$= \begin{cases} \int\cdots\int \rho_1(t_1)\ldots\rho_1(t_{k-1})\,\rho_1(t_{k+1})\ldots\rho_1(t_{p_1-1})\,\rho_2(t_{p_1})\ldots \\ \ldots\rho_n(t_{p+\ldots+p_n-1})\; dW(t_1)\ldots dW(t_{k-1})\; dW(t_{k+1})\ldots dW(t_p), & 1 \leqslant k \leqslant p_1-1, \\ 0 & k \geqslant p_1, \end{cases} \tag{6.11.8}$$

$$= \begin{cases} \prod_{i=2}^{n} \frac{H_{p_i}\left[\frac{1}{\sqrt{2}}\int \rho_1(t)\, dW(t)\right]}{(\sqrt{2})^{p_i}} \cdot \frac{H_{p_1-2}\left[\frac{1}{\sqrt{2}}\int \rho_1(t)\, dW(t)\right]}{(\sqrt{2})^{p_1-2,}} & 1 \leqslant k \leqslant p_1-1, \\ 0, & k \geqslant p_1. \end{cases}$$

We can now evaluate (6.11.4); the l.h.s. of (6.11.4) is the l.h.s. of (6.11.1) for the case $p + 1$, which we are trying to find, and the r.h.s. of (6.11.4) is:

$$\left[\prod_{i=2}^{n} \frac{H_{p_i}\left[\frac{1}{\sqrt{2}} \int \rho_i(t)\, dW(t)\right]}{(\sqrt{2})^{p_i}} \right] \cdot$$

$$\frac{1}{(\sqrt{2})^{p_1}} \left[\sqrt{2}\, H_{p_1-1}\left[\frac{1}{\sqrt{2}} \int \rho_1(t) dW(t)\right] \cdot \int \rho_1(t) dW(t) - 2(p_1-1)\, H_{p_1-2}\left[\frac{1}{\sqrt{2}} \int \rho_1(t) dW(t)\right]\right]. \quad (6.11.9)$$

The factor $(p_1 - 1)$ arises from the sum $\sum_k$ in (6.11.4) and the fact that (6.11.8) depends on k only through the given inequalities.

Now, since the Hermite polynomials satisfy a recursion relation:

$$H_{p_1}\left[\frac{x}{\sqrt{2}}\right] = \sqrt{2}x\, H_{p_1-1}\left[\frac{x}{\sqrt{2}}\right] - 2(p_1-1)\, H_{p_1-2}\left[\frac{x}{\sqrt{2}}\right], \quad (6.11.10)$$

(see table 5.2.1b), we see that the quantity within the right pair of brackets in (6.11.9) is simply

$$H_{p_1}\left[\frac{1}{\sqrt{2}} \int \rho_1(t)\, dW(t)\right].$$

Q.E.D.

6.12 Theorem: Completeness of Multiple Ito Integrals

Let $F\,[dW]$ be a lossy finite mean square functional of Wiener increments dW (see 5.5). Then, $F\,[dW]$ admits a multiple Ito integral expansion

$$F\,[dW] = \sum I_p\,(f_p) = \sum I_p\,(\tilde{f}_p), \quad (6.12.1)$$

the series converging in the mean-square sense.

Proof: By the Cameron-Martin theorem (interpreted on $(-\infty, \infty)$ as discussed in section 5.5),

$$F\,[dW] = \sum_{p} \sum_{p_1+\ldots+p_n=p} \sum_{u_1<\ldots<u_n} a_{p_1 \cdots p_n}^{u_1 \cdots u_n} \cdot$$

$$\prod_{i=1}^{n} H_{p_i} \left(\frac{1}{\sqrt{2}} \int \alpha_{u_i}(t)\,dW(t)\right), \tag{6.12.2}$$

where $\{\alpha_{u_i}(t)\}$ is a C.O.N. set, $\{H_i\}$ are the Hermite polynomials described in table 5.2.1b, and the a's are Fourier-Hermite coefficients related to the A's in (5.3.5) as follows. $H_p(x)$ is related to $H_p^N(x)$ by:

$$(2^p p!)^{-1/2} H_p \left(\frac{x}{\sqrt{2}}\right) = H_n^c \left(\frac{x}{\sqrt{2}}\right) = H_n^N(x), \tag{6.12.3}$$

therefore the product terms is simply a rescaled element of the Fourier-Hermite set $\{\Psi\}$ as in (5.3.3). The indexing on the summations is arranged in such a way that each unique product of terms appears only once. Therefore, the a's in (6.12.2) are related to the A's in (5.3.5) simply by the substitution of the Hermite polynomial is in (6.12.3), and proper bookkeeping of the indices:

$$a_{p_1 \cdots p_n}^{u_1 \cdots u_n} = A_{\cdots p_1 \cdots p_n} \prod_{i=1}^{n} (2^{p_i} p_i!)^{-1/2}, \tag{6.12.4}$$

where p_1 appears as a subscript of A in the u_1-th position, p_n appears as a subscript of A in the u_n-th position, and all the other subscripts are zero.

Let

$$f_p(t_1, \ldots, t_p) = 2^{p/2} \sum_{p_1+\ldots+p_n} \sum_{u_1<\ldots<u_n} a_{p_1 \cdots p_n}^{u_1 \cdots u_n} \alpha_{u_1}(t_1) \ldots \alpha_{u_1}(t_{p_1}) \ldots$$

$$\alpha_{u_n}(t_{p_1+\ldots+p_{n-1}+1}) \ldots \alpha_{u_n}(t_{p_1+\ldots+p_n}). \tag{6.12.5}$$

Then, by (6.11.1), (6.4.1), and (6.5.1), we obtain (6.12.1).

Q.E.D.

Remarks:

The assignment (6.12.5) yields a uniquely divided kernel f_p for each functional F. This is by virtue of the careful ordering of the subscripts, the symmetry of groupings of α's with the same subscript, the uniqueness of the A's in (5.3.5), and the corresponding uniqueness of the a's in (6.12.4). Thus, we have one proof that the symmetrized kernels $\tilde{f}_p$ in (6.12.1) are uniquely defined. In the next section, a more direct proof is given of this same result.

6.13 Theorem: Uniqueness of Symmetrized Kernels

$$\xi = \sum I_p(f_p) = \sum I_p(g_p) \Longrightarrow \tilde{f}_p = \tilde{g}_p \text{ a.e.} \tag{6.13.1}$$

Proof:

Let $\tilde{L}_2(T^p)$ be the totality of symmetric functions in $L_2(T^p)$. $\tilde{L}_2(T^p)$ is a closed subspace in $L_2(T^p)$. Let

$$F_p(\tilde{h}_p) = \frac{1}{p!} E\{\xi I^*_p(\tilde{h}_p)\}, \quad \tilde{h}_p \in \tilde{L}_2(T^p). \tag{6.13.2}$$

Since

$$F_p(a\tilde{h}_p + b\tilde{g}_p) = aF_p(\tilde{h}_p) + bF_p(\tilde{g}_p) \tag{6.13.3}$$

and

$$|F_p(\tilde{h}_p)| \leqslant \frac{1}{p!} ||\xi|| \cdot ||I_p(\tilde{h}_p)|| = ||\xi|| \cdot ||\tilde{h}_p||, \tag{6.13.4}$$

F_p is a bounded linear functional on $\tilde{L}_2(T^p)$. By Riesz-Fischer Theorem in Hilbert space, there exists $\tilde{s}_p \in \tilde{L}_2(T^p)$ such that

$$F_p(\tilde{h}_p) = \langle \tilde{s}_p, \tilde{h}_p \rangle . \tag{6.13.5}$$

By theorem 6.12,

$$F_p(\tilde{h}_p) = \frac{1}{p!} E\{I_p(\tilde{f}_p)\, I^*_p(\tilde{h}_p)\} = <\tilde{f}_p, \tilde{h}_p> . \tag{6.13.6}$$

Thus,

$$<\tilde{s}_p, \tilde{h}_p> = <\tilde{f}_p, \tilde{h}_p> \text{ for all } \tilde{h}_p \in \tilde{L}_2(T^p) . \tag{6.13.7}$$

This means that $\tilde{s}_p = \tilde{f}_p$ a.e.

Q.E.D.

"Lo! thy dread empire Chaos! is restor'd" ... Alexander Pope

7. Volterra and Wiener G-Functionals of White Noise

In this section, the multiple Ito integral is realized as a finite, telescoping series of homogeneous Volterra functionals driven by white noise. This series is called the Wiener G-functional. The G-functionals are an orthonormal set of functionals which, despite their appearance of being non-homogeneous, in fact are homogeneous over white noise excitations. they realize, in a subtractive manner, the masking of the diagonals inherent in the multiple Ito integral. The diagonal values of the G-functional kernels cannot be identified by the statistics of the G-functional, and may be assigned arbitrarily. Like the multiple Ito integrals, the G-functionals fully span the finite mean-square functionals of white noise, and can represent both non-casual and casual functionals of the white noise in series form.

Wiener G-functionals are part of the lore of non-linear systems analysis. Wiener G-functionals and Volterra series form the heart of the so-called Wiener-Volterra approach to identifying non-linear systems through the use of white noise excitation [RUG1, SHE1-2] which we will in fact use as part of our model realization procedure. This concept, attributable to Wiener [WIE2], is regrettably plagued with serious difficulties in general circumstances. Not every nonlinear system admits a Wiener G-functional series representation. The system

$$y(t) = (u(t))^2 \tag{7.1.1}$$

cannot be represented by Volterra series without resort to singularity function kernels, and it is pathologic under white noise excitation. (What *is* the square of a white noise, anyway?) Even when a non-linear system posesses a well-behaved Wiener series expansion under white noise excitation, the corresponding Volterra series description of the system may be disappointing. For example, the system,

$$y(t) = x(t)\cdot(1+x(t)^2)^{-1} \tag{7.1.2}$$

where x(t) is a stable linear functional of the input u(t), can be seen to admit virtually any practical input, including white noise. Unfortunately, the Volterra series resulting from Wiener series identification procedures is the Taylor expansion:

$$y(t) = x(t) - x(t)^3+x(t)^5-\ldots, \tag{7.1.3}$$

which is only valid for inputs defined by the bound $|x(t)|<1$. Polynomial series approximation is simply not a good choice for describing the response of a compressive system, such as this example, to all inputs.

In contrast to all of this, the Wiener-Volterra systems approach actually works very well for our purposes. The only systems we will identify are finite mean-square functionals of white noise which are known to admit a Wiener expansion. The Volterra systems of interest (actually, G functionals) will be finite order and will only be excited with standard white noise for the exclusive purpose of modelling general finite mean-square processes. Thus, while the Wiener-Volterra approach may deserve legitimate criticism ([PAL] for example) for being troublesome in general circumstances, it is truly in its element in our application.

7.2 Definition: Homogeneous Volterra Functional

Let $f_p(t_1,\ldots,t_p) \in L_2(T^p)$. Then f_p admits an orthonormal expansion in the L_2 sense as follows:

$$f_p(t_1,\ldots,t_p) = \lim_{N\to\infty} \sum_{\alpha_1,\ldots,\alpha_p}^{N} a_{\alpha_1\ldots\alpha_p}\phi_{\alpha_1}(t_1)\ldots\phi_{\alpha_p}(t_p),$$

where $\{\phi_{\alpha_i}\}$ is a C.O.N. set on T. Let

$$f_p^N(t_1,\ldots,t_p) = \sum_{\alpha_1,\ldots,\alpha_p}^{N} a_{\alpha_1\ldots\alpha_p}\phi_{\alpha_1}(t_1)\ldots\phi_{\alpha_p}(t_p)\,. \tag{7.2.2}$$

Then

$$||fp-fp^N||\rightarrow 0 \text{ as } N\rightarrow\infty\,. \tag{7.2.3}$$

Let u(t) $\in L^2(T)$ be a *deterministic* input function. Then, the p-th order homogeneous Volterra functional of u (with kernel f_p^N) is:

$$H_p(f_p^N, u) = \int_T\cdots\int_T f_p^N(t_1,\ldots,t_p)u(t_1)\ldots u(t_p)dt_1\ldots dt_p \tag{7.2.4}$$

$$\triangleq \sum_{\alpha_1,\ldots,\alpha_p}^{N} a_{\alpha_1\ldots\alpha_p}\prod_{i=1}^{p}\int_T\phi_{\alpha_i}(t_i)u(t_i)dt_i\,, \tag{7.2.5}$$

and

$$H_p(f_p,u) = \lim_{N\rightarrow\infty} H_p(f_p^N,u). \tag{7.2.6}$$

All of this is well defined because $f_p, f_p^N \in L_2(T^p)$ and u(t)$\epsilon L_2(T)$.

Let $X_n(t)\rightarrow n(t)$ be a sequence of mean-square continous Gaussian processes converging to white noise on T (see 3.2). Then the p-th order homogeneous Volterra functional of white noise n, Kernel f_p^N, is:

$$H_p(f_p^N,n) = \int_T\cdots\int_T f_p^N(t_1,\ldots,t_p)n(t_1)\ldots n(t_p)dt_1\ldots dt_p \tag{7.2.7}$$

$$= \sum_{\alpha_1,\ldots,\alpha_{p-1}}^{N} a_{\alpha_1\ldots\alpha_p}\prod_{i=1}^{p}\lim_{n\rightarrow\infty} \underline{q.m.}\int_T \phi_{\alpha_i}(t_i)X_n(t_i)dt_i \tag{7.2.8}$$

$$= \sum_{\alpha_1,\ldots,\alpha_p}^{N} a_{\alpha_1\ldots\alpha_p}\prod_{i=1}^{p}\int_T\phi_{\alpha_1}(t_i)n(t_i)dt_i \tag{7.2.9}$$

$$= S_o^{\frac{1}{2}}\sum_{\alpha_1,\ldots,\alpha_p}^{N} a_{\alpha_1\ldots\alpha_p}\prod_{i=1}^{p}\int_T\phi_{\alpha_1}(t_i)dW(t_i). \tag{7.2.10}$$

It should be noted that no attempt is made to directly define $n(t_1)\ \ldots n(t_p)$ as the limit of $X_n(t_1)\ldots X_n(t_p)$. The definition uses only the sense of $X_n(t)\rightarrow n(t)$ as given in section 3.2 (the linear sense). $H_p(f_p^N, n)$ is defined as a continous polynomic function (finite sum of finite products) of p finite mean-square *Gaussian linear functionals*

$$\int \phi_{\alpha_i}(t) X_n(t)\,dt \rightarrow \int \phi_{\alpha_i}(t) n(t)\,dt \tag{7.2.11}$$

of Gaussian processes $X_n(t) \rightarrow n(t)$.

7.3 Theorem: Linearity of H_p

$$H_p(uf_p + vg_p, n) = uH_p(f_p, n) + vH_p(g_p, n). \tag{7.3.1}$$

7.4 Theorem: Symmetrization

$$H_p(f_p) = H_p(\tilde{f}_p). \tag{7.4.1}$$

Proof: Let $(k_1, \ldots, k_p)$ be a permutation of $(\alpha_1, \ldots, \alpha_p)$, signified by $(k) \sim (\alpha)$, and let

$$V_{\alpha_i} = \int_T \phi_{\alpha_i}(t) n(t)\,dt. \tag{7.4.2}$$

Then

$$\prod_{i=1}^{p} V_{k_i} = \prod_{i=1}^{p} V_{\alpha_i}, \qquad (k) \sim (\alpha), \tag{7.4.3}$$

so

$$H_p(f_p, n) = \sum_{\alpha_1, \cdots, \alpha_p} a_{\alpha_1 \cdots \alpha_p} \prod_{i=1}^{p} V_{\alpha_i} \tag{7.4.4}$$

$$= \sum_{\alpha_1, \cdots, \alpha_p} a_{k_1 \ldots k_p} \prod_{i=1}^{p} V_{\alpha_i} \tag{7.4.5}$$

$$= \frac{1}{p!} \sum_{(k) \sim (\alpha)} \sum_{\alpha_1, \cdots, \alpha_p} a_{k_1 \ldots k_p} \prod_{i=1}^{p} V_{\alpha_i} \tag{7.4.6}$$

$$= \sum_{\alpha_1, \cdots, \alpha_p} \frac{1}{p!} \sum_{(k) \sim (\alpha)} a_{k_1 \ldots k_p} \prod_{i=1}^{p} V_{\alpha_i} \tag{7.4.7}$$

$$= H_p\,(\tilde{f}_{p,n}). \tag{7.4.8}$$

Q.E.D.

7.5 Theorem: Mean of $\mathbf{H_p}$

$\mathbf{E}\,\{H_p\,(f_{p,n})\} =$

$$\begin{cases} 0, & p \text{ odd}, \\ S_0^{\frac{p}{2}} \dfrac{p!}{\left(\frac{p}{2}\right)!2^{\frac{p}{2}}} \int \cdots \int \tilde{f}_p(t_1,t_1,\ldots,t_{\frac{p}{2}},t_{\frac{p}{2}})dt_1\ldots dt_{\frac{p}{2}}, & p \text{ even}. \end{cases} \tag{7.5.1}$$

Proof :

$$\begin{aligned} \mathbf{E}\{H_p(f_p,n)\} &= \mathbf{E}\{H_p(\tilde{f}_p,n)\} \\ &= \sum_{\alpha_1,\ldots,\alpha_p} \tilde{a}_{\alpha_1\ldots\alpha_p}\,\mathbf{E}\left\{\prod_{i=1}^{p} V_{\alpha_i}\right\}. \end{aligned} \tag{7.5.2}$$

Using (5.11), we see the r.h.s. of (7.5.2) is zero for p odd. Using (5.12), for p even, the r.h.s of (7.5.2) is

$$\sum_{\alpha_1,\ldots,\alpha_p} \tilde{a}_{\alpha_1\ldots\alpha_p}\, S_0^{\frac{p}{2}} \sum\prod \delta_{\alpha_i\alpha_j}. \tag{7.5.3}$$

There are

$$\frac{p!}{(\frac{p}{2})!2^{\frac{p}{2}}} \tag{7.5.4}$$

terms in $\sum\prod$, each one selecting a particular pairing of indices α_i,α_j. but since $\tilde{a}$ is symmetric, we get

$$S_0^{\frac{p}{2}} \frac{p!}{(\frac{p}{2})!2^{\frac{p}{2}}} \sum_{u_1,\ldots u_{\frac{p}{2}}} \tilde{a}_{u_1u_1\ldots u_{\frac{p}{2}}u_{\frac{p}{2}}}. \tag{7.5.5}$$

Using (7.2.2) and the orthonomality of $\{\phi_{\alpha_i}\}$, we evaluate the sum in (7.5.5) as an equivalent integral along diagonals:

$$S_0^{\frac{p}{2}} \frac{p!}{(\frac{p}{2})!2^{\frac{p}{2}}} \int \cdots \int \tilde{f}(t_1,t_1,\ldots t_{\frac{p}{2}},t_{\frac{p}{2}})dt_1\ldots dt_{\frac{p}{2}}. \tag{7.5.6}$$

Q.E.D.

7.6 Theorem: Mean Square of $\mathbf{H_p}$ is Finite

$$\mathbf{E}|H_p(f_p,n)|^2 < \infty. \tag{7.6.1}$$

Proof : Let $f_p(t_1, \ldots, t_p() \in L_2(T^p)$ be elementary. Let

$$y = H_p(f_p,n). \tag{7.6.2}$$

Then

$$|y|^2 = \int \cdots \int f_p(t_1, \ldots, t_p) f_p^*(t_{p+1}, \ldots, t_{2p}) n(t_1) \ldots n(t_{2p}) dt \ldots dt_{2p} \tag{7.6.3}$$

$$= H_{2p}(g_{2p},n), \tag{7.6.4}$$

where

$$g_{2p}(t_1, \ldots, t_{2p}) = f_p(t_1, \ldots, t_p) f_p^*(t_{p+1}, \ldots, t_{2p}). \tag{7.6.5}$$

Since

$$\int \cdots \int |g_{2p}|^2 dt^{2p} = \left[\int \cdots \int |f_p|^2 dt^p\right]^2 < \infty, \tag{7.6.6}$$

we see that $g_{2p} \in L_2(T^{2p})$. It is also elementary, so indeed (7.6.4) is well defined.

By Theorem 7.5,

$$\mathbf{E}\{|y|^2\} = E\{H_{2p}(g_{2p},n)\} \tag{7.6.7}$$

$$\leqslant \frac{(2p)!}{p!2^p} \int \cdots \int |\tilde{g}(t_1,t_1, \ldots, t_p,t_p)| dt_1 \ldots dt_p \tag{7.6.8}$$

$$< \infty\,. \tag{7.6.9}$$

Q.E.D.

7 .7 Definition: $\mathbf{H_p(f_p,n)}$ for general $\mathbf{f_p}$.

Theorems 7.6 and 7.3 prove that H_p is a bounded, linear functional on the elementary kernels f_p^N. Therefore H_p can be extended to all $f \in L_2(T^p)$ as follows: Let $f_p^N \rightarrow f_p$. Then

$$H_p(f_p,n) \triangleq \lim_{n \to \infty} q.m.\ (f_p^N,n).$$

With this definition, Theorems 7.3-7.6 also extend to $H_p(f_p,n)$.

7.8 Definition: Wiener G-Functional

Let H_p denote the p-th order homogeneous Volterra functional operator on white noise $n(t)$:

$$H_p(h_p,n) = \int \cdots \int h_p(t_1, \ldots, tp) n(t_1) \cdots n(t_p) dt_1 \ldots dt_p, \qquad (7.8.1)$$

and adopt the convention

$$H_0(h_0,n) = h_0. \qquad (7.8.2)$$

Let $n(t)$ have spectral density S_0, and let $\tilde{h}_p$ denote line symmetrization of p-th order kernel $h_p \in L_2(T^p)$.

The p-th order Wiener G-functional G_p is:

$$G_p(k_p,n) \triangleq \sum_{i=0}^{p} H_i(k_{i(p)},n) \qquad (7.8.3)$$

where the *derived kernels* $k_{i(p)}$ (derived from k_p) are defined as follows:

$$k_{p(p)}(t_1, , \ldots, t_p) = \tilde{k}_p(t_1, \ldots, t_p), \qquad (7.8.4)$$

$$k_{p-2m(p)}(tt_1, \ldots, t_{p-2m})$$
$$= \frac{(-1)^m p! S_0^m}{(p-2m)!(m)!2^m} \int \cdots \int \tilde{k}_p(x_1,x_1, \ldots, x_m,x_m,t_1, \ldots, t_{p-2m}) dx_1 \ldots dx_m,$$
$$m = 1, \ldots, \left[\frac{p}{2}\right], \qquad (7.8.5)$$

$$k_{i(p)} = 0, \quad i \neq p - 2m \text{ for some } m = 0, \ldots, \left[\frac{p}{2}\right]. \qquad (7.8.6)$$

Remarks: The G-functional G_p is frequently called a non-homogeneous p-th order Volterra functional because it appears as a telescoping series of homogeneous Volterra functionals of decreasing order with kernels of similarly decreasing order. It should be noted, however, that S_0 *is actually a functional of* $n(t)$ (albeit a constant). If this functional dependence of S_0 on $n(t)$ is respected, *the G-functional is indeed a homogeneous* *functional of white noise inputs n(t)*. This will be demonstrated by showing the relationship of G_p to I_p, the homogenous multiple Ito integral. In essence, the auxulliary, derived integrals of order p-2m are precisely the components thrown away in summation $\sum''$ in the development of the multiple Ito integral. Along the diagonals in $\sum''$, the square increments $S_0 dW^2$ act (in the m.s. sense) like $S_0 dt$. With S_0 a sure functional of $n(t)$, we can then, in effect, precompute the integral of $S_0 dW^2$ along diagonals as sure inner product integrals with respect to $S_0 dt$ along those same diagonals.

It is trivial to show that the G-functional is linear with respect to its kernels:

$$G_p(uf_p + vg_p, n) = uG_p(f_p, n) + vG_p(g_p, n) \qquad (7.8.7)$$

and that

$$G_p(f_p, n) = G_p(\tilde{f}_p, n). \qquad (7.8.8)$$

In [SHE_1] and [RUG_1], the G-functionals are introduced as a potential means for identifying a deterministic Volterra system. In these accounts, the G-functionals are defined by requiring the orthogonality condition:

$$\mathbf{E}\{H_m(h_m, n) G_p(k_p, n)\} = 0, \qquad m < p \qquad (7.8.9)$$

for any m-th order homogeneous Volterra functional H_m. The kernels (7.8.4) through (7.8.6) are the ones that achieve this condition. Thus, in these works, (7.8.9) is used to *define* the G functionals, and (7.8.4) through (7.8.6) are the derived results.

In this paper the selected role for the G-functionals is a white noise driven Volterra realization of the multiple Ito integral I_p. In this context, the approach in [SHE_1] and [RUG_1] is somewhat more indirect, even though it is perfectly sound. I simply choose to show how the Volterra Series follows from the multiple Ito series rather than the other way around. With so much machinery already in place in Sections 5 and 6, the development here of the G-functional properties from the adopted definition 7.2 and 7.8 and the recursion theorem 7.9 is very direct and rapid.

7.9 Theorem: Recursion Relation

Let $\{\phi_i(t)\}$ be an orthonormal set of real-valued functions on $L_2(T)$. Let

$$\Phi(t_1, \ldots, t_p) = \phi_{i_1}(t_1) \ldots \phi_{i_p}(t_p). \tag{7.9.1}$$

Then the G functionals of white noise $n(t)$, spectral density S_0, satisfy the recursion relation:

$$G_{p+1}\left[\Phi\phi_{i_{p+1}}\right] = G_p(\Phi)\, G_1\left[\phi_{i_{p+1}}\right] - S_0 \sum_{k=1}^{p} G_{p-1}\left[\Phi \underset{(k)}{\times} \phi_{i_{p+1}}\right] \tag{7.9.2}$$

where

$$\Phi \underset{(k)}{\times} \phi_{i_{p+1}}\left[t_1, \ldots, t_{p-1}\right] = \int \Phi(t_1, \ldots, t_{k-1}, t_k, \ldots, t_p)\phi_{i_{p+1}}(t_k)\,dt_k. \tag{7.9.3}$$

Proof: All three terms in (7.9.2) comprise integrals that are well-defined. The theorem will be proven by showing a direct relation between each functional and kernel on the l.h.s. of (7.9.2) of order $p+1-2m$ and the functionals and kernels in the r.h.s. of (7.9.2) of the same order. This will be done by splitting each $(p+1-2m)$-th order term on the l.h.s. into two groups which contribute the respective parts to the same order terms on the r.h.s.

Let

$$k_{p+1}(t_1, \ldots, t_{p+1}) = \phi_{i_1}(t_1) \ldots \phi_{i_{p+1}}(t_{p+1}) \tag{7.9.4}$$

and let $\tilde{k}_{p+1}$ denote its symmetrization. Then we write the derived kernels of k_{p+1} as:

$$k_{p+1-2m(p+1)}(t_1, \ldots, t_{p+1-2m}) =$$
$$\frac{(-1)^m S_0^m}{(p+1-2m)!(m)!2^m} \sum_{(j)\sim(i)} \int \ldots \int \phi_{j_1}(x_1)\phi_{j_2}(x_1) \ldots \phi_{j_{2m-1}}(x_m)\phi_{j_{2m}}(x_m)$$
$$\ldots \phi_{j_{2m+1}}(t_1) \ldots \phi_{j_{p+1}}(t_{p+1-2m})\,dx_1 \ldots dx_m, \tag{7.9.5}$$

where the sum is over all permutations $(j_1, \ldots, j_{p+1})$ of $(i_1, \ldots, i_{p+1})$, and where $m = 1, \ldots, \left[\frac{p+1}{2}\right]$.

Many of the $(p+1)!$ terms in the sum are identical, since:

- $\phi_{i_j}(x)\phi_{i_k}(x) = \phi_{i_k}(x)\phi_{i_j}(x)$ (2^m repeats),

- commuting *pairs* of functions ϕ within the integral does not change the integral ($m!$ repeats),

- and permutations of the $p+1-2m$ functions ϕ outside the integral does not change the product ($(p+1-2m)!$ repeats).

We now break the sum in (7.9.5) into two groups:

$$\sum_{(j)\sim(i)} \ldots = \Sigma_1 \ldots + \Sigma_2 \ldots, \tag{7.9.6}$$

corresponding to Groups 1 and 2:

- Group 1: $\phi_{i_{p+1}}$ does not occur within the integral, i.e., it is *free*, $(p+1-2m)p!$ terms, and

- Group 2: $\phi_{i_{p+1}}$ is inside the integral, is *integrated out*, 2 mp! terms.

In Group 1, permulations of $\phi_{i_{p+1}}$ with the other free $\phi's$ has no effect on the multiple white noise integral, and this is exploited by merely noting the number of permutations and by factoring out a first-order functional with kernel $\phi_{i_{p+1}}$ from the remainder. In Group 2, the terms can be ordered according the the subscript of the ϕ that combines with $\phi_{i_{p+1}}$ in the integral. By making this ordering explicit, we obtain the Group 2 terms in a desirable form.

We first consider the case when p is even. The lead kernel $(m=0)$ is written as follows:

$$k_{p+1(p+1)}\left(t_1, \ldots, t_{p+1}\right) = \frac{1}{(p+1)!} \sum_{(j)\sim(i)} \phi_{j_1}(t_1)\ldots\phi_{j_{p+1}}(t_{p+1}). \tag{7.9.7}$$

All occurrences of $\phi_{i_{p+1}}$ in the sum are free, so we put all these terms in Group 1, and nothing in Group 2. Moreover,

$$H_{p+1}(k_{p+1(p+1)}, n) = H_{p+1}\left(\frac{p+1}{(p+1)!}\phi_{i_{p+1}}(t_{p+1}) \sum_{(j)\sim(i)} \phi_{j_1}(t_1)\ldots\phi_{j_p}(t_p),\ n\right) \tag{7.9.8}$$

$$= H_p(k_{p(p)}, n)\, G_1(\phi_{i_{p+1}}, n), \tag{7.9.9}$$

where

$$k_{p(p)} = \frac{1}{p!}\sum_{(j)\sim(i)} \Phi\left(t_{j_1}, \ldots, t_{j_p}\right) \tag{7.9.10}$$

$$= \tilde{k}_p, \tag{7.9.11}$$

and

$$k_p(t_1, \ldots, t_p) = \Phi(t_1, \ldots, t_p). \tag{7.9.12}$$

Now, for $m = 1, \ldots, \frac{p}{2}$, and working with Σ_1 first,

$$H_{p+1-2m}\left(\frac{(-1)^m S_0^m}{(p+1-2m)!m!2^m}\Sigma_1 \ldots,\ n\right) =$$

$$H_{p+1-2m}\left(\frac{(-1)^m S_o^m}{(p+1-2m)!m!2^m}(p+1-2m)\phi_{i_{p+1}}(t_{p+1-2m}) \cdot \right.$$

$$\left.\sum_{(j)\sim(i)} \int\cdots\int \phi_{j_1}(x_1)\phi_{j_2}(x_1)\ldots dx_1\ldots dx_m,\ n\right) \tag{7.9.13}$$

$$= H_{p-2m}(k_{p-2m(p)}, n)\, G_1\left(\phi_{i_{p+1}}, n\right), \tag{7.9.14}$$

where

$$k_{p-2m(p)}(t_1, \ldots, t_{p-2m})$$
$$= \frac{(-1)^m S_o^m}{(p-2m)!m!2^m} \sum_{(j)\sim(i)} \int \cdots \int \phi_{j_1}(x_1)\phi_{j_2}(x_1) \ldots$$
$$\phi_{j_{2m-1}}(x_m)\phi_{j_{2m}}(x_m)\phi_{j_{2m+1}}(t_1) \ldots \phi_{j_p}(t_{p-2m})\,dx_1 \ldots dx_m. \qquad (7.9.15)$$

This is the formula for $k_{p-2m(p)}$ as derived from $k_p = \Phi$, and is the result we seek.

Next, we work with Σ_2:

$$\Sigma_2 \ldots = 2m \sum_{k=1}^{p} \sum_{(j)\sim(i) \neq k} \int \cdots \int \phi_{j_i}(x_1)\phi_{j_2}(x_1) \ldots$$
$$\phi_{i_k}(x_k)\phi_{i_{p+1}}(x_k) \ldots \phi_{j_{2m}}(x_m)\,dx_1 \ldots dx_m \phi_{j_{2m+1}} \ldots \phi_{j_p}(t_{p+1-2m}). \qquad (7.9.16)$$

The notation $(j)\sim(i) \neq k$ denotes the $(p-1)!$, permutations of all subscripts except k and $p+1$. The sum Σ_2 has been expressed as two nested sums, the outer one freezing the ϕ_{i_k} with which $\phi_{i_{p+1}}$ is paired. The factor $2m$ accounts for the invariance of the integrals to interchange of ϕ_{i_k} with $\phi_{i_{p+1}}$ and the m possible permutations of the pair $(\phi_{i_k}, \phi_{i_p+1})$ with the other ϕ pairs within the integral.

We may rewrite (7.9.16) as:

$$\Sigma_2 \ldots = 2m \sum_{k=1}^{P} \sum_{(j)\sim(i)} \Phi_p \underset{(k)}{\times} \phi_{i_{p+1}}, \qquad (7.9.17)$$

and therefore,

$$\frac{(-1)^m S_0^m}{(p+1-2m)!m!2^m} \sum_2 \cdots \qquad (7.9.18)$$
$$= \frac{(-1)^m S_0^m 2m}{(p+1-2m)!m!2^m} \sum_{k=1}^{p} \sum_{(j)\sim(i)} \Phi_p \underset{(k)}{\times} \phi_{i_{p+1}}$$
$$= -S_0 \sum_{k=1}^{p} \frac{(-1)^{m-1} S_0^{m-1}}{(p-1-2(m-1))!(m-1)!2^{m-1}} \sum_{(j)\sim(i)} \Phi_p \underset{(k)}{\times} \phi_{i_{p+1}} \qquad (7.9.19)$$
$$= -S_0 \sum_{k=1}^{p} k^k_{p-1-2(m-1),(p-1)} \; (t_1, \ldots, t_{p-1-2(m-1)}), \qquad (7.9.20)$$

where $k^k(\ldots)$ is recognized as the kernel derived from $\Phi_p \underset{(k)}{\times} \phi_{i_{p+1}}$ for the case $m-1=0,\ldots,\left[\frac{p-1}{2}\right]$. This is the desired result for Σ_2, and proves the theorem for p even. Note that, for $m = 1, \ldots \frac{p}{2}$, there are terms appearing in both Σ_1 and Σ_2, owing to $p+1$ being odd. That is, over all $(p+1)!$ permutations, $\phi_{i_{p+1}}$ appears both free and within integrals for every $m=1,\ldots,\frac{p}{2}$.

Now, for p odd. The difference between this case and p even is tied to the allowed ranges for m in each of the sums Σ_1 and Σ_2. As in (7.9.7), when $m=0$ on the l.h.s. of (7.9.2), all terms have $\phi_{i_{p+1}}$ free, and therefore all terms are accounted for in Σ_1. The cases $m=1,\ldots,\frac{p-1}{2}$ also follows the previous development for p even, since, for these values of m , there is always a mix of free occurrences of $\phi_{i_{p+1}}$ and occurrences within the integrals.

For $m = \frac{p+1}{2}$, there are no free occurrences of $\phi_{i_{p+1}}$. This is the case that gives rise to the constant term in the G-functional series. The formulas given for Σ_1 in (7.9.13) though (7.9.15) understandably do not make sense for this value of m. However, all required terms in Σ_1 are already accounted for by $m=0,\ldots,\frac{p-1}{2}$. The formulas for Σ_2 in (7.9.17) through (7.9.20) are valid for $m-1 = \left[\frac{p-1}{2}\right]$. Thus, the constant terms associated with the $(p-1)$-th order G_{p-1} functionals are properly defined and accounted for, and their sum (over index k) equals the constant term associated with the $(p+1)$-th order functional G_{p+1}.

Q.E.D.

7.10 Theorem: Relationship between G-Functional and Fourier-Hermite Expansions.

Let $\{\phi_i(t)\}$ be a C.O.N. set of real-valued functions in $L_2(T)$, let $n(t)$ be white noise, spectral density S_0, and let $\tilde{H}_p(x)$ be the version of p-th order Hermite polynomial in table 5.2.1e. Then, with $p = p_1 + \ldots + p_n$,

$$G_p\,(\phi_1(t_1)\ \ldots\ \phi_1(t_{p_1})\ \phi_2(t_{p_1+1}) \ldots \phi_2(t_{p_1+p_2}) \ldots \phi_n(t_{p_1+\ldots p_{n-1}+1}) \ldots \phi_n(t_p), n)$$

p_i term

$$= \prod_{i=1}^{n} \tilde{H}_{p_i}\left(\int \phi_i\,(t) n(t) dt\right). \tag{7.10.1}$$

Proof. Follows the same proof by induction as in Theorem 6.11. Some conversion of Hermite polynomials is necessary; for example:

$$\prod_{i=1}^{n} \frac{H_{p_i}\left(\frac{1}{\sqrt{2}}\int \phi_i(t) dW(t)\right)}{(\sqrt{2})^{p_i}} = \prod_{i=1}^{n} \tilde{H}_{p_i}\left(\int \phi_i(t) n(t) dt\right). \tag{7.10.2}$$

7.11 Definition: G functional, f ∈ $L_2(T^p)$.

If $f_p, f_p^N \in L_2(T^p)$ and

$$||f_p^N - f_p|| \to 0 \text{ as } N \to \infty, \tag{7.11.1}$$

Then

$$G_p(f_p, n) = \lim_{N\to\infty} \text{q.m } G_p(f_p^N, n). \tag{7.11.2}$$

7.12 Theorem: Relationship between G-Functional and Multiple Ito Integral

Let $k_p \in L_2(T^p)$, real-valued, and let $n(t)$ be white noise, spectral density S_0, and $W(t)$ the standard Wiener process corresponding to $n(t)$. Then

$$G_p(k_p,n) = S_0^{\frac{p}{2}} \; I_p\,(k_p, W). \tag{7.11.1}$$

Proof. Let $\{\phi_i(t)\}$ be a C.O.N set on $L_2(T)$. By Theorems 6.11 and 7.10, the result has already been proven for kernels of the form:

$$\phi_{i_1}(t_1) \; \ldots \; \phi_{i_p}\,(t_p)\,. \tag{7.11.2}$$

Approximate k_p by

$$k_p^N(t_1, \ldots, t_p) = \sum_{i_1, \ldots i_p}^{N} a_{i_1, \ldots, i_p}\, \phi_{i_1}(t_1) \; \ldots \; \phi_{i_p}(t_p), \tag{7.11.3}$$

so that

$$||k_p - k_p^N|| \to 0 \text{ as } N \to \infty. \tag{7.11.4}$$

Then

$$G_p(k_p, n) = \lim_{N\to\infty} q.m. \; G_p(k_p^N, n) = \lim_{N\to\infty} q.m. \; S_0^{\frac{p}{2}} I_p\,(k_p^N, W)$$
$$= S_0^{\frac{p}{2}} \; I_p(k_p, W). \tag{7.11.5}$$

Q.E.D.

"I am as strong as a bull moose and you can use me to the limit." ... Theodore Roosevelt.

8. Dynamical Realizations for Stationary, Causal Wiener Series

The purpose of this section is to show methods by which Wiener Series can be realized by finite-dimensional dynamical systems driven by white noise. The hope is that, by finding at least one finite-dimensional realization with the desired properties, one can achieve reduced-order models using minimal-order realization concepts borrowed from state-space methods in control theory.

In section 10, I will give a *causal, stationary* functional of white noise with the desired approximation characteristics, and in section 9, I will show how the corresponding *causal, stationary* Wiener Series can be identified. In this section, therefore, I assume I am given a functional Y of white noise which has one of the following forms:

Fourier-Hermite expansion:

$$Y(t) = \sum_{m_1, \ldots, m_N} A_{m_1 \cdots m_N} \prod_m H_m^N \left(\int_{-\infty}^{t} l_m (t - \tau)\, dW(\tau) \right), \tag{8.0.1}$$

multiple Ito integral expansion:

$$Y(t) = \sum_n \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} h_n (t - \tau_1, \ldots, t - \tau_n)\, dW(\tau_1) \ldots dW(\tau_n), \tag{8.0.2}$$

or Wiener G-functional expansion:

$$Y(t) = \sum_n \left[\sum_{i=0}^{n} \int_{-\infty}^{t} \int_{-\infty}^{t} h_{i(n)} (t - \tau_1, \ldots, t - \tau_n)\, n(\tau_1) \ldots n(\tau_n)\, d\tau_1 \ldots d\tau_n \right], \tag{8.0.3}$$

where either the A's in (8.0.1) or the h_n's in (8.02) and (8.03) are given. A particular goal is to show is how Y may be computed via a bilinear stationary stochastic differential equation:

$$\begin{aligned} d\mathbf{X}_t &= \mathbf{A}\mathbf{X}_t\, dt + \mathbf{N}\mathbf{X}_t\, dW_t + \mathbf{B}\, dW_t \\ dY_t &= \mathbf{C}\, d\mathbf{X}_t, \end{aligned} \tag{8.0.4}$$

and, how this, in turn, can be realized by a bilinear stationary differential equation driven by white noise:

$$\frac{d}{dt} \mathbf{X}(t) = \mathbf{A}' \mathbf{X}(t) + \mathbf{N}'\mathbf{X}(t)\, n(t) + \mathbf{B}'n(t) + \mathbf{D}'. \qquad (8.0.5)$$
$$Y(t) = \mathbf{C}'\mathbf{X}(T)$$

8.1 Truncating The Wiener Series

Recall from section 5 that $Y(t)$ can be approximated in the mean-square sense by a *finite* sum of the form (8.0.1). This series is *finite* in two ways: there is a maximum *finite order* of Hermite polynomials, and there is a *finite number* of basis functions l_m, such as the Laguerre functions (5.5.2). There is, therefore, a justification for truncating the Wiener Series, the result being a finite series with kernels which are *finite* sums of *finite* products:

$$h_n\,(t_1, \ldots, t_n) = \sum_{i_1 \cdots i_n}^{N} a_{i_1 \ldots i_n}\, l_{i_1}(t_1) \ldots l_{i_n}(t_n) \qquad (8.1.1)$$

8.2 Definition: A kernel h_n is *separable* if it can be written as a finite sum of products:

$$h_n\,(t_1, \ldots, t_n) = \sum_{i_1 \cdots i_n}^{N} v_{i_1}(t_1) \ldots v_{i_n}(t_n)\,. \qquad (8.2.1)$$

The kernel h_n is said to be *differentially separable* if each of the v's above is differentiable.

By truncating the Wiener Series, and adopting the Laguerre functions l_i for the v's, we obtain a finite order series each of which is readily identified and has kernels which are differentably separable.

8.3 Theorem: [BRO1, RUB] A finite Volterra Series

$$Y\,(t) = \sum_{l=0}^{N} \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} h_i(t - t_1, \ldots, t - t_i)\; u(t_1) \ldots u(t_i)\, dt_1 \ldots dt_i\,, \qquad (8.3.1)$$

where u is continuous on $(-\infty, t)$, is realizable via a finite-dimensional bi-linear differential equation of the form:

$$\frac{d}{dt}\mathbf{X}(t) = \mathbf{A}(t)\mathbf{X}(t) + \mathbf{N}(t)\mathbf{X}(t)u(t) + \mathbf{B}(t)u(t) \tag{8.3.2}$$
$$Y(t) = \mathbf{C}(t)\mathbf{X}(t)$$

if and only if the kernels h_i, $i = 0, \ldots, N$ are separable. The bilinear realization can be made stationary (time-invariant) if and only if the kernels are differentiably separable.

The details of this proof may be found in the references. The proofs show how separability allows each kernel h_n to be written in the triangular form:

$$h_n\,(t - t_1, \ldots, t - t_n) = \sum_{i_1 \cdots i_n}^{N} v_{i_1}(t - t_1)\, v_{i_2}(t_1 - t_2) \ldots v_{i_n}(t_{n-1} - t_n). \tag{8.3.3}$$

Note that the Volterra integrals can be evaluated over triangular regions:

$$\int_{-\infty}^{t} \cdots \int_{-\infty}^{t} h_i(t-t_1, \ldots, t-t_i)u(t_1) \ldots u(t_i)\, dt_1 \ldots dt_i$$
$$= i! \int_{-\infty}^{t} \int_{-\infty}^{t_{i-1}} \cdots \int_{-\infty}^{t_2} \tilde{h}_i\,(t - t_1, \ldots, t - t_i)\, u(t_1) \ldots u(t_i)\, dt_1 \ldots dt_i, \tag{8.3.4}$$

where $\tilde{h}$ denotes the symmetrization of h. If (8.3.3) holds, then (8.3.4) can be evaluated by iterated integrals. Each iteration integrates in a linear fashion the instantaneous product of input $u(t)$ with the previous integration result. When the kernels are differentiably separable, these successive integrations can be done by time-invariant systems. If one writes each step or iteration, the model (8.0.4) becomes quickly apparent. Associated with each integration is a block, and the matrix N describes how the output of a given block is multiplied by the input and fed to the next block.

An example should help illustrate this procedure. Note that the Laguerre functions can be written as sums of functions of the form $t^p e^{\lambda t}$. Consider realizing

$$\int_{-\infty}^{t} \int_{-\infty}^{t} (t - t_1)^2\, e^{\lambda_1(t-t_1)}\, (t - t_2)^2\, e^{\lambda_2(t-t_2)}\, u(t_1)u(t_2)\,dt_1 dt_2 \tag{8.3.5}$$

as a stationary, bilinear system. Note that

$$e^{\lambda_1(t-t_1)}\, e^{\lambda_2(t-t_2)} = e^{(\lambda_1+\lambda_2)(t-t_2)}\, e^{\lambda_2(t_1-t_2)}, \tag{8.3.6}$$

and that

$$\begin{aligned}(t-t_1)^2(t-t_2)^2 &= (t-t_1)^2(t-t_1+t_1-t_2)^2 \\ &= (t-t_1)^4 + 2(t-t_1)^3(t_1-t_2) + (t-t_1)^2(t_1-t_2)^2.\end{aligned} \tag{8.3.7}$$

Then, (8.3.5) can be evaluated as the sum of two groups of integrals, one being:

$$\begin{aligned}&\int_{-\infty}^{t} (t-t_1)^4\, e^{(\lambda_1+\lambda_2)(t-t_1)}\, u(t_1) \left[\int_{-\infty}^{t_1} e^{\lambda_2(t_1-t_2)}\, u(t_2)\, dt_2 \right] dt_1 \\ &+ \int_{-\infty}^{t} (t-t_1)^3\, e^{(\lambda_1+\lambda_2)(t-t_1)}\, u(t_1) \left[\int_{-\infty}^{t_1} (t_1-t_2)\, e^{\lambda_2(t_1-t_2)}\, u(t_2)\, dt_2 \right] dt_1 \\ &+ \int_{-\infty}^{t} (t-t_1)^2\, e^{(\lambda_1+\lambda_2)(t-t_1)}\, u(t_1) \left[\int_{-\infty}^{t_1} (t_1-t_2)^2\, e^{\lambda_2(t_1-t_2)}\, u(t_2)\, dt_2 \right] dt_1.\end{aligned} \tag{8.3.8}$$

The other has the indices 1 and 2 interchanged and takes into account the symmetrization. By taking the bracketted integral first, then the outer integral second, it should become clear that (8.3.7) can be realized by a stationary, bilinear system.

Note that the only condition imposed on u here is that it be continuous. Normally, in matters of Volterra series, it is necessary to further stipulate bounds on u to assure convergence. What is making the difference here is that the truncated series is of *finite order*. So, truncation not only leads to finite realizations, but also realizations which exhibit broad tolerances to various excitations.

Note in the example above that, to achieve the bilinear realization, it was necessary to symmetrize the kernel. This is a necessary step that greatly increases the dimension of the required state space, perhaps even beyond that for an equivalent Fourier-Hermite expansion. This is an example of a *bilinear* system requiring greater state dimension than a *linear-analytic* one (see [BRO1]).

8.4 Theorem

A truncated causal stationary Wiener G-functional series can be realized as a finite order, stationary bilinear system driven by white noise.

Proof: Recall a white noise is a sequence of almost surely *continuous* Gaussian processes $X_n(t) \rightarrow n(t)$. The lead Volterra operators are realizable as bilinear systems for each path of $X_n(t)$. Each of the derived operators is also a Volterra operator with differentially separable kernels (products of the Laguerre functions integrate out nicely into lower-order products of Laguerre functions) and so is realizable as a bilinear system.

For every continuous path of $X_n(t)$, both the original G-functional of $X_n(t)$ and the bilinear realization of the G-functional of $X_n(t)$ yield identical solutions. The response of both the G-functional and its bilinear realization precisely match as $X_n(t) \rightarrow n(t)$, and so by the definition of the G-functional of white noise, the theorem is proved.

Note that additional states would be required in order to realize the derived kernels. We will now show how the G-functionals can be realized using only those states needed to realize the lead operator.

8.5 Definition: Ito Stochastic Integral [WON]

Let W_t, $a \leqslant t \leqslant b$, be a Wiener process. Correspondingly, there is an increasing family of sub-σ algebras of A on which W_t is defined, such that for each s, $\{W_t - W_s, t \geqslant s\}$ is independent of A_s and W_t is A_t measurable for each t.

Let $f(\omega, t)$ be a function jointly measurable in (ω, t) such that, for each t,

$$f_t = f(\cdot, t) \text{ is measurable with respect to } A_t, \text{ and} \tag{8.5.1}$$

$$\int_a^b \mathbf{E}\, |f_t|^2 \, dt < \infty . \tag{8.5.2}$$

If f is an (ω, t) step function:

$$f(\omega, t) = f_k(\omega), \qquad t_k \leqslant t \leqslant t_{k+1}, \tag{8.5.3}$$

where $a = t_0 < t_1 < \cdots < t_n = b$ and $k = 0, \cdots, n-1$, then the Ito Stochastic integral of f with respect to increments dW is given by:

$$\int_a^b f(\omega, t)\, dW(\omega, t) = \sum_{k=0}^{n-1} f_k(\omega)\, [\, W(\omega, t_{k+1}) - W(\omega, t_k)] \,. \tag{8.5.4}$$

If f is not necessarily a step function, but still satisfies (8.5.1) and (8.5.2), there exists a sequence of (ω, t) step functions $f_n(\omega, t)$ satisfying (8.5.1) and (8.2.2) such that

$$\| f - f_n \|^2 = \int_a^b \mathbf{E}\, |f(\cdot, t) - f_n(\cdot, t)|^2\, dt \to 0 \quad \text{as } n \to \infty \,. \tag{8.5.5}$$

The Ito stochastic integral of f with respect to dW is given by

$$\int_a^b f(\omega, t)\, dW(\omega, t) = \lim_{n \to \infty} \text{q.m.} \int_a^b f_n(\omega, t)\, dW(\omega, t). \tag{8.5.6}$$

8.6 Theorem

Let $f(t_1, \ldots, t_p) \in L_2(T^p)$, $T = [a, t]$. Then the multiple Wiener-Ito integral $I_p(f_p)$ from section 6 can be evaluated as iterated stochastic integrals:

$$\begin{aligned} I_p(f_p) &= p! \int_a^t \left\{ \int_a^{t_p} \left(\cdots \int_a^{t_3} \left\{ \int_a^{t_2} \tilde{f}(t_1, \ldots, t_p)\, dW_{t_1} \right\} dW_{t_2} \cdots \right\} dW_{t_{p-1}} \right\} dW_{t_p} \\ &= \sum_{(\pi)} \int_a^t \left(\cdots\cdots \int_a^{t_2} f\left(t_{\pi_1}, \ldots, t_{\pi_p} \right) dW_{t_1} \right) \cdots\cdots \Bigg) dW_{t_p} \,. \end{aligned} \tag{8.6.1}$$

Proof. Recall $I_p(f_p) = I_p(\tilde{f}_p)$. By symmetrizing $f_p \to \tilde{f}_p$, we create a $p!$-fold symmetry, of T^p that is fully represented by the wedge:

$$\begin{aligned} a &\leq t_p \leq t \\ a &\leq t_{p-1} \leq t_p \\ &\;\;\vdots \\ a &\leq t_1 \leq t_2 \end{aligned} \tag{8.6.2}$$

Recall $f \in L_2(T^p)$ can be approximated in the mean-square sense by *special* functions which are dense in $L_2(T^p)$. For each such function:

$$f = \prod_{i=1}^{p} \chi_{T_i}, \qquad T_i \cap T_j = \varnothing, \tag{8.6.3}$$

the multiple Wiener-Ito integral is simply

$$I_p(f) = \prod_{i=1}^{p} \beta(T_i). \tag{8.6.4}$$

But this is precisely the same result we obtain using the iterated stochastic Ito integral as in (8.5.4). So (8.6.1) holds for f_p a special function.

If f_p is not a special function, let $f_{s,p} \overset{\text{m.s.}}{\longrightarrow} f_p$, where $f_{s,p}$ is a special function. By (8.5.5) and (8.5.6), again the iterated Ito integral precisely agrees with the multiple Wiener-Ito integral.

Remark: The result is easily extended to the case $a \to -\infty$.

8.7 Theorem

Suppose the p-th order Volterra operator on continuous u:

$$\int_0^t \cdots \int_0^t h_p(t_1, \ldots, t_p)\; u(t_1) \ldots u(t_p) dt_1 \ldots dt_p \tag{8.7.1}$$

is realizable as a stationary, finite dimensional bilinear system:

$$\begin{aligned} \frac{d}{dt}\mathbf{X}(t) &= \mathbf{A}\mathbf{X}(t) + \mathbf{N}\mathbf{X}(t)u(t) + \mathbf{B}u(t) \\ Y(t) &= \mathbf{C}\mathbf{X}(t). \end{aligned} \tag{8.7.2}$$

Then the multiple Wiener-Ito integral

$$\int_0^t \cdots \int_0^t h_p(t_1, \ldots, t_p)\; dW_{t_1} \ldots dW_{t_p} \tag{8.7.3}$$

is realizable as a stationary, finite dimensional bilinear system:

$$dX_t = AX_t + NX_t dW_t + BdW_t \tag{8.7.4}$$
$$dY_t = CdX_t,$$

where the integration in (8.7.4) is in the Ito sense.

Proof. Recall that (8.7.1) is realizable as a stationary, finite dimensional bilinear system if and only if h_p is differentiably separable. Recall this implies (8.7.1) can be written as:

$$\int_0^t v_p(t-t_p) \int_0^{t_p} \cdots \int_0^{t_2} v_1(t_2-t_1)u(t_1)dt_1 \ldots dt_p . \tag{8.7.5}$$

where

$$h_p(t-t_1, \ldots, t-t_p) = v_p(t-t_p) \cdots v_1(t_2-t_1) . \tag{8.7.6}$$

By theorem 8.6, (8.7.3) can therefore be written as

$$\int_0^t v_p(t-t_p) \int_0^{t_p} \cdots \int_0^{t_2} v_1(t_2-t_1)dW_{t_1} \cdots dW_{t_p} . \tag{8.7.7}$$

If (8.7.2) corresponds (block) row for row, with the evaluation of a nested integral, the result (8.7.4) is immediate. If not, there exists rotation of states (see [RUB]) that will result in this correspondence for both (8.7.2) and (8.7.4).

8.8 Theorem: [WON] Wong Zakai Correction

Let

$$\frac{d}{dt}X(t) = m(X(t), t) + \sigma(X(t), t)\,\xi(t) \tag{8.8.1}$$

be a given stochastic differential equation such that $m(X,t)$, $\sigma(X,t)$, $\frac{\partial}{\partial X_i}\sigma(X,t)$, $i = 1, \ldots, n$, and $\frac{\partial}{\partial t}\sigma(X,t)$ are continuous on $-\infty < X_i < \infty$, $a \leqslant t \leqslant b$. Also, let m, σ, and $\sigma\frac{\partial \sigma}{\partial X_i}$ satisfy a uniform Lipschitz condition:

$$\| f(X,t) - f(Y,t) \| \leqslant K \| X - Y \|. \tag{8.8.2}$$

where f stands for $\mathbf{m}$, σ, and $\sigma \frac{\partial \sigma}{\partial X_i}$ one at a time. Let $|| \sigma || \geqslant K_1 > 0$ and $||\sigma|| < K_2 || \sigma ||^2$. If $\xi \rightarrow n(t)$ a white noise with $S_0 = 1$, and $dW(t)$ is the Wiener increments process associated with $n(t)$, then the solution $\mathbf{X}$ converges almost surely on $a \leqslant t \leqslant b$ to the solution of:

$$d\mathbf{X}_t = \mathbf{m}\ (\mathbf{X}_t, t)\ dt + \frac{1}{2} \sum_{l,m} \frac{\partial\ \sigma_{km}}{\partial X_l}\ (\mathbf{X},t)\ \sigma_{lm}\ (\mathbf{X},t) + \boldsymbol{\sigma}\ (\mathbf{X}_t, t)\ dW_t\,. \quad (8.8.3)$$

8.9 Theorem: Application to Stationary, Bilinear Equations

$$\text{Let}\quad \frac{d}{dt}\mathbf{X}\ (t) = \mathbf{A}\mathbf{X}(t) + \mathbf{N}\ \mathbf{X}(t)\ \xi + \mathbf{B}\ \xi\ (t) + \mathbf{D}\,. \qquad (8.9.1)$$

If $\xi \rightarrow n(t)$, with $S_0 = 1$, then $\mathbf{X}$ converges to the solution of:

$$d\ \mathbf{X}_t = (\mathbf{A}\,\mathbf{X}_t + \frac{1}{2}\ \mathbf{N}\mathbf{N}\mathbf{X}_t + \frac{1}{2}\ \mathbf{N}\mathbf{B} + \mathbf{D})\ dt + \mathbf{N}\mathbf{X}_t\ dW_t + \mathbf{B}dW_t. \qquad (8.9.2)$$

8.10 Theorem:

The Ito differential equation

$$d\mathbf{X}_t = \mathbf{A}\mathbf{X}_t\ dt + \mathbf{N}\mathbf{X}_t dW_t + \mathbf{B}dW_t \qquad (8.10.1)$$

Can be realized via the white-noise-driven differential equation:

$$\frac{d}{dt} = \mathbf{A}'\ \mathbf{X}_t + \mathbf{N}\ \mathbf{X}_t\ n(t) + \mathbf{B}n(t) + \mathbf{D} \qquad (8.10.2)$$

where $n(t)$ is the white noise associated with $dW(t)$, with $S_0 = 1$,

and

$$\mathbf{A}' = \mathbf{A} - \frac{1}{2}\ \mathbf{N}\mathbf{N}, \qquad (8.10.3)$$

$$\mathbf{D} = -\frac{1}{2}\ \mathbf{N}\mathbf{B}\,.$$

The proof of 8.9 follows from 8.8 quickly by realizing that stationary bilinear differential equations satisfy the conditions in 8.8. The proof of 8.10 follows from 8.8 and 8.9 by simply adopting (8.10.2) and showing (8.10.1) is the corresponding Ito equation.

Theorem 8.10 shows that, if you find a bilinear realization of the lead Volterra operator of a G-functional you can realize the entire G-functional (or multiple Wiener -Ito integral) without needing to separately realize the *derived* operators. You simply make two corrections, one to the A matrix, the other a fixed bias.

Chaos often breeds life . . .
Henry Brooks Adams

9. Identification of Wiener Series Kernels

In this section, I describe cross-correlation methods for determining the kernels of the Wiener expansion of a given L_2 functional of white noise. These methods are often referred to as Lee-Shetzen identification (after two of Wiener's colleagues, Y. W. Lee and Martin Shetzer, who popularized many of Wiener's ideas). They are usually introduced within the context of non-linear system identification using white noise as the excitation, the so-called Wiener-Voltarra approach to non-linear system analysis [SHE1, RUG1]. As discussed in section 7, there are problems with applying Lee-Shetzen methods indiscriminately. Fortunately, however, our application is well suited to this identification procedure. The systems we will identify will always be L_2 functionals of white noise. The Wiener Series developed will be of finite order and will always be excited by white noise for the purpose of realizing stochastic processes.

Two methods are described, both of which use cross-correlations or projection techniques. The first, or direct, method, attempts to measure the p-th order kernels directly by cross-correlation of the functional output which p-th order products of delayed, filtered samples of the white noise input. In my description, I make a deliberate attempt to avoid taking products of white noise itself or using delta functions. (Compare with the accounts in [SHE1, SHE2, and RUG1].) This is done to clarify the inherent limitation of the method in resolving the kernel along diagonals. It also emphasizes that the kernels, which belong to $L_2(T^p)$, can only be resolved in an averaged, mean-square sense and not truly point-wise as delta function descriptions might lead one to believe. The second, or Fourier-Hermite, method is based on the material in Section 4.

9.1 "Direct" Method

Let $F(n(t))$ be an L_2 functional of white noise $n(t)$, and let its Wiener expansion be given by

$$F(n) = \sum_{i=0}^{\infty} G_p(h_p, n) \tag{9.1.1}$$

where G_p is given by (7.8.3), (7.8.4), (7.8.5). Construct the following smoothed functional of $n(t)$:

$$q_\epsilon(\tau) = S_0^{-1/2} \frac{1}{\epsilon} \int_{\tau-\epsilon}^{\tau} n(\xi)\, d\xi \, . \tag{9.1.2}$$

Choose $\tau_1, \ldots, \tau_p$ such that the intervals $[\tau_i - \epsilon, \tau_i]$ are all disjoint, and let

$$Q_{p,\epsilon}(\tau_1, \ldots, \tau_p) = \prod_{i=1}^{p} q_\epsilon(\tau_i) \, . \tag{9.1.3}$$

Then $Q_{p,\epsilon}$ is a Wiener G-functional of $n(t)$ of order p. (Note how diagonals are avoided).

Let the estimated kernel $h_{p,\epsilon}$ be defined by:

$$h_{p,\epsilon}(\tau_1, \ldots, \tau_p) = \frac{1}{p!} \mathrm{E} \left\{ F(n)\, Q_{p,\epsilon}(\tau_1, \ldots, \tau_p)(n) \right\} \tag{9.1.4}$$

By (7.11.5), (6.6.1) and (6.7.1),

$$\begin{aligned} h_{p,\epsilon}(\tau_1, \ldots, \tau_p) &= \frac{1}{p!} \mathrm{E} \left\{ G_p(h_p, n)\, Q_{p,\epsilon}(\tau_1, \ldots, \tau_p)(n) \right\} \\ &= \frac{1}{\epsilon^p} \int_{\tau_1-\epsilon}^{\tau_1} \cdots \int_{\tau_p-\epsilon}^{\tau_p} h_p(t_1, \ldots, t_p)\, dt_1 \ldots dtp \, . \end{aligned} \tag{9.1.5}$$

We resolve how $h_{p,\epsilon}$ behaves as $\epsilon \to 0$ as follows.

9.2 Theorem

Let $h_p \in L_2(T^p)$ and $h_{p,\epsilon}$ defined as above. Then, for any $g_p \in L_2(T^p)$,

$$\lim_{\epsilon \to 0} \mathbf{E} \left\{ \left| G_p(g_p,n) \left[G_p(h_{p,\epsilon},n) - G_p(h_p,n) \right] \right|^2 \right\} = 0. \tag{9.2.1}$$

Proof: This follows from (6.8.2) if, for all $g_p \in L_2(T^p)$,

$$\int \cdots \int \left| g_p \, (h_{p,\epsilon} - h_p) \right|^2 dt^p \to 0 \text{ as } \epsilon \to 0. \tag{9.2.2}$$

Now, if h^p is continuous, (9.2.1) is quickly obtained using step functions as the g's. If h_p is not continuous, it can be approximated by a continuous function $\tilde{h}_p$, such that

$$\int \cdots \int \left| g_p \, (h_{p,\epsilon} - h_p) \right|^2 dt^p \leqslant \int \cdots \int \left| g_p \, (\tilde{h}_{p,\epsilon} - \tilde{h}_p) \right|^2 dt^p + \delta(\epsilon), \tag{9.2.3}$$

where $\tilde{h}_{p,\epsilon}$ is estimated from $\tilde{h}_p$ and δ is made arbitrarily small.

9.3 Remarks

Note that the τ_i must be distinct for $Q_{p,\epsilon}$ to be a p-th order Wiener functional. As $\epsilon \to 0$, the τ's may be made arbitrarily close together, but no two τ_i may be made equal. Recall that the *Ito* integrals I_p and Wiener functionals G_p do not depend on the kernel diagonals. We may therefore assign diagonal values for $h_{p,\epsilon}$ arbitrarily, and so this restriction on the τ's is not a problem for us.

The behavior of $Q_{p,\epsilon}$ for overlapping τ can be interpreted as follows. Suppose $\tau_1 = \tau_2$. Then

$$Q_{p,\epsilon} = \frac{1}{\epsilon^2} \frac{1}{S_0} \left[\int_{\tau_1 - \epsilon}^{\tau_1} n(\xi) \, d\xi \right]^2 \prod_{i=3}^{p} q_\epsilon(\tau_i) \tag{9.3.1}$$

$$\approx \frac{1}{\epsilon} \prod_{i=3}^{p} q_\epsilon(\tau_i) \text{ as } \epsilon \to 0. \tag{9.3.2}$$

Thus, $Q_{p,\epsilon}$ actually behaves like a functional of order less than p, and therefore fails to identify useful information about kernel h_p in direction of $G_p(h_p, n)$.

Theorem 9.2 shows that, while $h_{p,\epsilon}$ may not converge pointwise to h_p, it is clearly equivalent to h_p as a substitute within a Wiener G-functional. (Of course, if h_p is smooth, we *do* obtain pointwise convergence!)

Finally, it should be noted that the only critical property of filter q_ϵ in (9.1.2) is that it have a band pass that converges to unity over the frequency spectrum supporting the functional F. This is because $L_2(T)$ approximations to q_ϵ result in $L_2(T^p)$ approximations to $Q_{p,\epsilon}$. The particular form of q_ϵ adopted in (9.1.2) simply assures without fuss that, for each ϵ, the corresponding functional $Q_{p,\epsilon}$ is clearly a homogenous p-th order Wiener G-functional.

9.4 Fourier-Hermite Method

This method is simply the implementation in steps (5.3.2) through (5.3.5) in the Cameron-Martin Theorem. This results in the determination of the Fourier-Hermite coefficients of the Fourier-Hermite series expansion, in contrast to a Wiener Integral expansion.

"Sing unto him a new song;

Play skillfully with a loud noise."

... Psalm 33

10. Wiener-Nisio Functional

Thus far, the only convergence idea introduced has been the mean-square convergence of Wiener Series expansions of a given L_2 functional of white noise. We now explore a broader concept: convergence (in a sense to be described) of the *process* generated by Wiener Series under the flow of the white noise, specifically toward given processes which are to be modelled.

Let $\{X(\omega^*,t), -\infty < t < \infty\}$ be a given stochastic process, defined on some probability space $(\Omega^*,\mathbf{B}^*,\mathbf{P}^*)$, having the following properties:

$$\begin{aligned} &\text{a. scalar, real-valued,}\\ &\text{b. strictly stationary,}\\ &\text{c. ergodic,}\\ &\text{d. continuous in probability.} \end{aligned} \tag{10.0.1}$$

Let $X(\cdot,\omega_0^*)$ denote a typical sample path of X. (This is a representative sample of data.) Using this path, we shall construct a sequence of causal finite mean-square functionals of an arbitrarily selected white noise excitation. Under the flow of the white noise, these functionals give rise to a sequence of processes that converge in finite-order distribution to X. Each of these functionals can be approximated by finite-order, finite-dimensional causal Wiener Series. By this means, we demonstrate constructively that may process satisfying (10.0.1) can be approximated in the sense of finite distributions by a finite-order causal dynamical system driven by white noise.

The Wiener-Nisio functional involves a remarkable construct, a sampling time determined by paths of a white noise or Wiener increments process. The idea of the random sampling time is apparently due to Wiener [WIE1]. The concept was refined and enlarged by Nisio [NIS1], and was summarized by McKean [MCK1]. Nisio proves the approximation theorem both with and without the ergodic assumption (10.0.1c), but the latter is not pursued here—intentionally. The ergodic version emphasizes how a typical sample path can serve as the sufficient statistic for modelling purposes without the necessity of explicitly determining process moments or other statistical characterizations.

The original versions of the Wiener-Nisio functional described in [NIS1] and [MCK1] are almost causal (from the point of view of building causal realizations). I will demonstrate how a modification yields causal functionals. I will also discuss the relationship of convergence in the sense of Nisio's U-topology to that of weak convergence in an effort to point out a fundamental limitation inherent in any realization procedure based on actual data: the data may not support statistical match beyond certain limits, such as finite distributions.

10.1 Random Sampling Time

Let $n(\omega,t)$ and $W(\omega,t)$ be a standard white noise and a corresponding Wiener increments process defined on $-\infty < t < \infty$. Let ω_s^+ denote the shifted path of ω by s; that is $W(t,\omega_s^+) = W(t + s,\omega)$, so $s > 0$ implies a lead. We now define a sequence of random sampling times derived from n (and W), denoted $\lambda_n(\omega)$, as follows. Let

$$S(\omega) = \{t: | W(t+1,\omega) - W(t-1,\omega) | \} > 1. \tag{10.1.1}$$

$S(\omega)$ is open by virtue of the continuity of paths W, and can therefore be written as a countable disjoint union of open intervals $I_i(\omega)$. Let m denote Lebesque measure on $\boldsymbol{R}$, define

$$T_n(\omega) = \{I_i(\omega): m(I_i(\omega)) > n, \; I_i(\omega) \subset (-n,\infty)\}, \tag{10.1.2}$$

and let

$$S_n(\omega) = \bigcup_{I_i \in T_n} I_i(\omega). \tag{10.1.3}$$

Thus, S_n is the set of open intervals starting at $t = -n$ for which the Wiener increments in (10.1.1) exceed unity for the least n units of t. (See Figure 10.1.1).

We next show that $S_n(\omega)$ is not empty for every $n \geqslant 1$ and for almost all ω. Let $f(\omega)$ be any measurable function of ω with respect to the Borel field determined by the increments $\{W(t,\omega) - W(s,\omega), -\infty < s < t < \infty\}$. By the ergodicity of $W(t,\omega)$ and the stationarity of its increments, we have

$$\lim_{T\to\infty} \frac{1}{T} \int_a^{a+T} f(\omega_t^+)\, dt = \mathrm{E}\,\{f(\omega)\}. \tag{10.1.4}$$

Specifically let

$$f(\omega) = \begin{cases} 1, & |W(t+1) - W(t-1)| > 1, 0 \leqslant t \leqslant n \\ & \text{and } |W(-n+1) - W(-n-1)| \leqslant 1, \\ 0 & \text{otherwise.} \end{cases} \tag{10.1.5}$$

Since $f(\omega_t^+) = 1$ implies $t \in S_n(\omega)$, we have

$$\begin{aligned} &\liminf_{T\to\infty} \frac{1}{T}\, m(S_n(\omega) \cap (-n, -n+T)) \\ &\geqslant \lim \frac{1}{T} \int_{-n}^{-n+T} f(\omega_t^+)\, dt = \mathrm{E}\,\{f(\omega)\} \end{aligned} \tag{10.1.6}$$

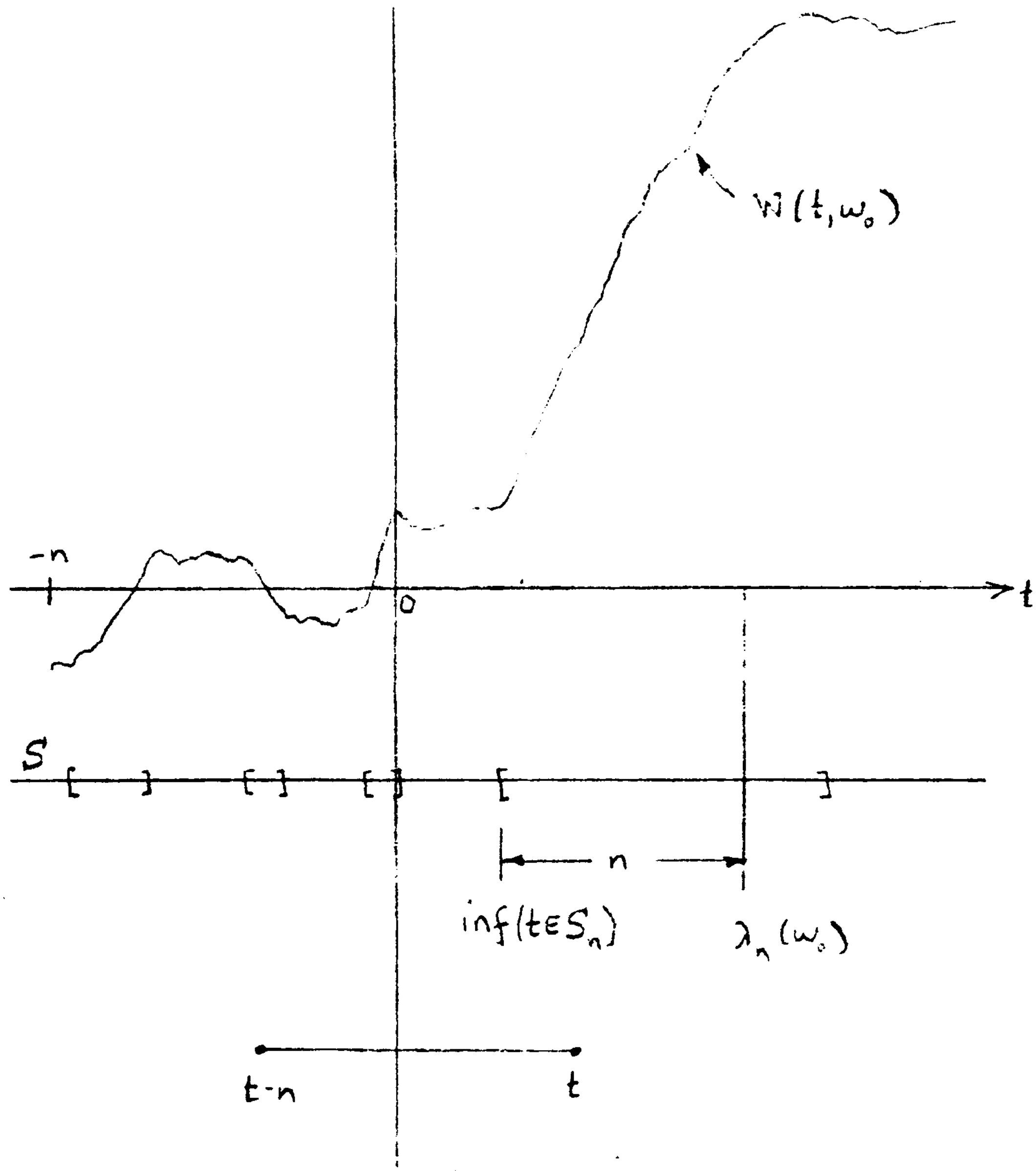


Fig. 10.1.1 — Illustration of terms defining $\lambda_n(\omega)$

$$= \mathbf{P}\{|W(t+1) - W(t-1)| > 1, 0 \leqslant t \leqslant n, \text{ and } |W(-n+1) - W(-n-1)| \leqslant 1\} \quad (10.1.7)$$

$$= \mathbf{P}\{|W(t+1) - W(t-1)| > 1, 0 \leqslant t \leqslant n\} \cdot \mathbf{P}\{|W(-n+1) - W(-n-1)| \leqslant 1\} \quad (10.1.8)$$

for almost all ω. The second probability is strictly >0 trivially, so it is enough to show

$$\mathbf{P}\{|W(t+2) - W(t)| > 1, 0 \leqslant t < n\} > 0. \quad (10.1.9)$$

But the ω-set in (10.1.9) includes the intersection of the following ω-sets:

$$\{2k-2 < W(t) - W(0) < 2k+1, k \leqslant t \leqslant k+1\}, \quad k = 0,1,\ldots,n-1 \quad (10.1.10)$$

(See figure 10.1.2). Note these sets adjoin along unit intervals, and so by the construction of the Wiener process, they have positive probability. This proves that (10.1.9) is true, and therefore that $S_n(\omega)$ is always non-empty.

Now we define a sequence of *random times* λ_n dependent on ω through dW as follows:

$$\lambda_n(\omega) = n + inf\,(t : t \in S_n(\omega)), \quad (10.1.11)$$

with $\lambda_n(\omega) = \infty$ if $S_n(\omega) = \varnothing$. By the above, $\lambda_n(\omega)$ is finite almost surely.

10.2 Theorem: $\lambda_n(\omega)$ is almost Uniformly Distributed on [0,n].

That is, $X_n(\omega)$ has a probability density which is zero on $(-\infty,0)$, is flat on [0,n], and decreases on (n, ∞). (See Figure 10.2.1.)

Proof: $\lambda_n(\omega) = t$ is equivalent to:

$$t - n \notin S(\omega),\ (t-n,t) \subset S(\omega) \text{ if } 0 \leqslant t \leqslant n,$$

and

$$t - n \notin S(\omega),\ (t-n,t] \subset S(\omega),\ (-n,t-n) \cap S_n(\omega) = \varnothing \text{ if } t > n$$

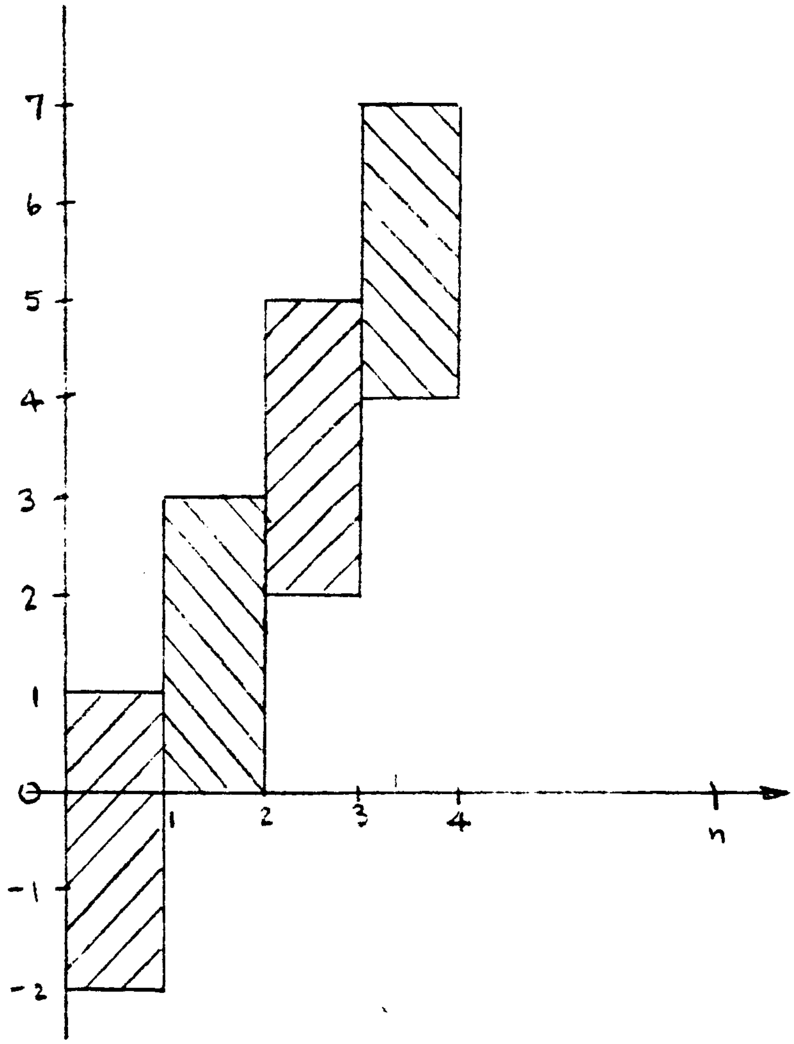


Fig. 10.1.2 — Illustration of sets defined by Eqn. 10.1.10

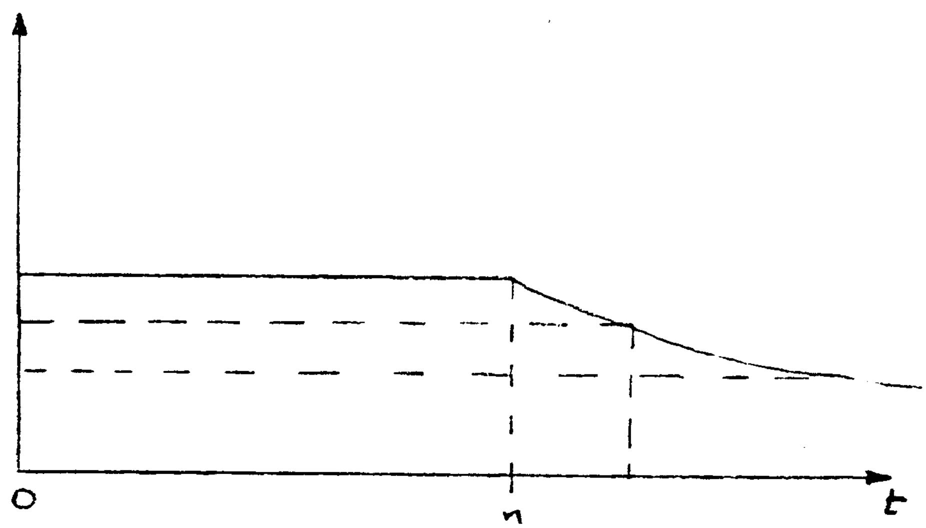


Fig. 10.2.1 — Probability density of λ_n

Note that if $0 \leqslant t \leqslant n$, then $t - n \notin S(\omega)$ implies that $(-n, t - n) \cap S_n(\omega) = \varnothing$

For $t_1 + s \leqslant n$, $s > 0$ and $t < 0$ (see Figure 10.2.2a),

$$
\begin{aligned}
&\mathbf{P}\{\lambda_n \in (t + s, t_1 + s)\} \\
&\quad = \mathbf{P}\{\exists x \in (t + s, b_1 + s) | x - n \notin S(\omega), (x - n, x) \subset S(\omega)\} \\
&\quad = \mathbf{P}\{\exists x \in (t, t_1) | x - n \notin S(\omega_s^+), (x - n, x] \cap S(\omega_s^+)\} \\
&\quad = \mathbf{P}\{\exists x \in (t, t_1) | x - n \notin S(\omega), (x - n, x] \subset S(\omega)\}
\end{aligned}
$$

by the stationarity of the increments of W. Thus,

$$\mathbf{P}\{\lambda_n \in (t + s, t_1 + s)\} = \mathbf{P}\{\lambda_n \in (t, t_1)\} . \tag{10.2.1}$$

Now, if $t_1 + s > n$, $s > 0$ and $t > 0$,

$$
\begin{aligned}
&\mathbf{P}\{\lambda_n \in (t + s, t_1 + s)\} \\
&\quad = \mathbf{P}\{\exists x \in (t + s, t_1 + s) | x - n \in S(\omega), (x - n, x] \subset S(\omega), \\
&\quad (-n, x - n) \cap S_n(\omega) = \varnothing\} \\
&\quad = \mathbf{P}\{\exists x \in (t, t_1) | x - n \in S(\omega_s^+), (x - n, x] \subset S(\omega_s^+), \\
&\quad (-n - s, x - n) \cap S_n(\omega_s^+) = \varnothing\} \\
&\quad \leqslant \mathbf{P}\{\exists x \in (t, t_1) | x - n \in S(\omega_s^+), (x - n, x] \subset S(\omega_s^+), \\
&\quad (-n, x - n) \cap S_n(\omega_s^+) = \varnothing\} \\
&\quad = \mathbf{P}\{\lambda_n \in (t, t_1)\}
\end{aligned} \tag{10.2.2}
$$

The result follows from (10.2.1) and (10.2.2).

Q.E.D.

10.3 Theorem: $\lambda(\omega)$ almost preserves shifts in dW(t).

That is, for any $l > 0$ and $\epsilon > 0$, there exists $N(l, \epsilon)$ such that, if $n > N$,

$$\mathbf{P}\{\lambda_n(\omega_s +) = \lambda_n(\omega) - s, -l \leqslant s \leqslant l\} > 1 - \epsilon. \tag{10.3.1}$$

Proof: The proof relies on expressing $\lambda_n(\omega_s^+) = \lambda_n(\omega) - s$ in a suitable form. First, let $s > 0$. Then, by the construction of $\lambda_n(\omega)$,

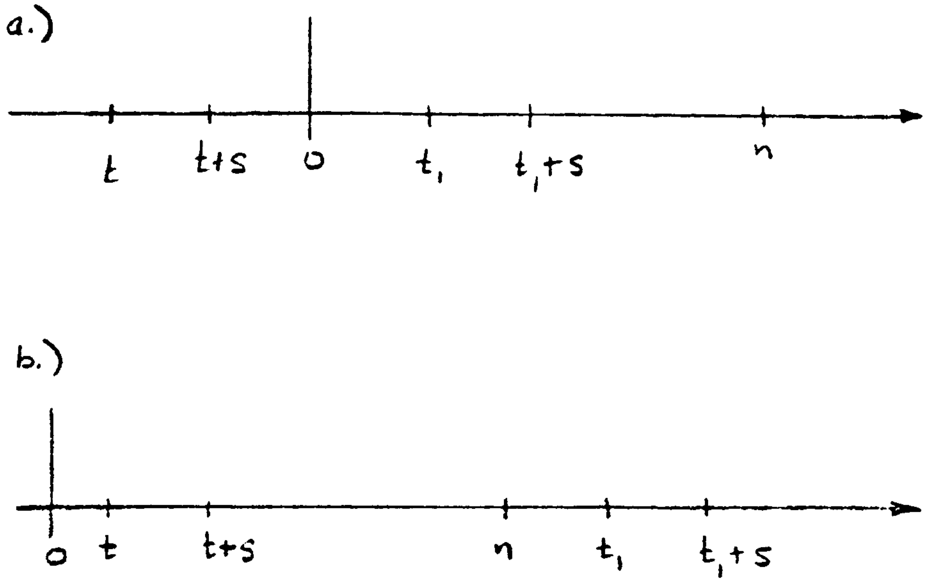


Fig. 10.2.2 — Ordering of variables on time line

$$\lambda_n(\omega_s+) = t$$
$$<==> t - Sn \notin S(\omega_s^+), (t-n,t] \subset S(\omega_s+), (-n,t-n) \cap S_n(\omega_s^+) = \varnothing$$
$$<==> t + s - n \in S(\omega), (t+s-n, t+s] \subset S(\omega),$$
$$(-n+s, t+s-n) \cap S_n(\omega) = \varnothing. \tag{10.3.2}$$

But

$$\lambda_n(\omega) = t + s$$
$$<==> t+s-n \in S(\omega), (t+s-n, t+s) \subset S(\omega), \tag{10.3.3}$$
$$(-n, t+s-n) \cap S_n(\omega) = \varnothing.$$

This means that

$$\lambda_n(\omega_s^{+)} = \lambda_n(\omega) - s \text{ iff } (-n, -n+s] \cap S_n(\omega) = \varnothing, \tag{10.3.4}$$

which is also equivalent to

$$\lambda_n(\omega) \geqslant s.$$

Therefore, if $s > s_1 > 0$ and if $\lambda_n(\omega_s^+) = \lambda_n(\omega) - s$, then

$$\lambda_n(\omega_{s_1}^+). = \lambda_n(\omega) - s_1$$

Next, if $s > 0$, let $r = -s$. Note that $\omega = (\omega_{-r}^+)_r^+$ and $\omega_s^+ = \omega_{-r}^+$. Applying the previous result to ω_{-u}^+, we find that

$$\lambda_n((\omega_{-r}^+)_r^+) = \lambda_n(\omega_{-r}^+) - r$$

is equivalent to

$$(-n, -n+r) \cap S_n(\omega_{-r}^+) = \varnothing, \text{ or to } \lambda_n(\omega_{-r}^+) \geqslant r.$$

If $s < s_1 < 0$ and if $\lambda_n(\omega_s^+) = \lambda_n(\omega) - s$, then $\lambda_n(\omega_{s_1}^+ = \lambda_n(\omega) - s_1$, and this is because $S_n(\omega_s^+) \supset S_n(\omega_{s_1}^+)$ which follows from $(S(\omega) - s) \cap (-n, \infty) \supset (S(\omega) - s_1) \cap (-n, \infty)$.

We now use these results to show the following:

$$\begin{aligned}&\mathbf{P}\{\lambda_n(\omega_s^+) = \lambda_n(\omega) - s, -l \leqslant s \leqslant l\}\\ &= \mathbf{P}\{\lambda_n(\omega_l^+) = \lambda_n(\omega) - l \text{ and } \lambda_n(\omega_{-l}^+) = \lambda_n(\omega) + l\}\\ &1 - \mathbf{P}\{\lambda_n(\omega_l^+) \neq \lambda_n(\omega) - l\} - \mathbf{P}\{\lambda_n(\omega_{-l}^+) \neq \lambda_n(\omega) + l\}\ .\end{aligned} \tag{10.3.6}$$

From theorem 10.2 and the stationarity of dW, we know that

$$\mathbf{P}\{\lambda_n(\omega_l^+) \leqslant l\} \leqslant \frac{l}{n}, \tag{10.3.7}$$

and

$$\mathbf{P}\{\lambda_n(\omega_{-l}^+) \leqslant l\} = \mathbf{P}\{\lambda_n(\omega) \leqslant l\} \leqslant \frac{l}{n}. \tag{10.3.8}$$

Therefore,

$$\mathbf{P}\{\lambda_n(\omega_s^+) = \lambda_n(\omega) - s, -l \leqslant s \leqslant l\} \geqslant 1 - \frac{2l}{n}.$$

Q.E.D.

10.4 Nisio's U and V Topologies

Definition: $U(Z,\epsilon)$ is the collection of all stochastic processes $\{Y(t,\omega),-t < \infty\}$ such that

$$|\mathbf{E}\ e^{j\theta X_1 + \cdots j\theta_n X_n} - \mathbf{E}\ e^{j\theta_1 Y_1 + \cdots + j\theta_n Y_n}| < \epsilon \tag{10.4.1}$$

whenever $n, |\theta_i|$ and $|t_i|$ are all less than $\frac{1}{\epsilon}$.

Definition: $V(X,\epsilon)$ is the collection of all stochastic processes y such that

$$\mathbf{P}\{|Y(t) - X(t)| > \epsilon\} < \epsilon \tag{10.4.2}$$

whenever $|t| < \frac{1}{\epsilon}$.

Remark: Note that convergence in the V topology implies convergence in the U topology. To show this, assume (10.4.2). Then

$$|\mathbf{E}\, e^{j\theta_1 X(t_1)+\ldots+j\theta_n X(t_n)} - \mathbf{E}\, e^{j\theta_1 Y(t_1)+\cdots+j\theta_n Y(t_n)}|$$
$$= |\mathbf{E}\left[e^{j[\theta_1(X(t_1)-Y(t_1)]+\cdots+j[\theta_n(X(t_n)-Y(t_n))]}\right] - 1|$$
$$\leqslant 2\epsilon + \prod_{i=1}^{n} 4\sin^2\frac{\theta_i\epsilon}{2} \leqslant 2\epsilon + \theta_{\max}^2\epsilon^2. \tag{10.4.3}$$

So, if $\theta_{\max} < n$, by making $\epsilon < \frac{1}{3n^2}$ we hold

$$|\mathbf{E}\, e^{j\cdots}| \leqslant \frac{2}{3n^2} + \frac{n^2}{9n4} = \frac{7}{9n^2} \tag{10.4.4}$$

which proves that

$$V(X,\frac{1}{3n^2}) \subset (X,\frac{1}{n}). \tag{10.4.5}$$

Also note that in both V and U topologies, if Y belongs to the ϵ_1 neighborhood of X, and Z belongs to the ϵ_2 neighborhood Y, then Z belongs to the $(\epsilon_1 + \epsilon_2)$ neighborhood of X.

10.5 Development of the Wiener-Nisio Functional

Let $\omega^* \in \Omega^*$ and let $(\Omega^*, \mathbf{B}^*, \mathbf{P}^*)$ be the underlying probability space. Let $\{X(t,\omega^*) - \infty < t < \infty\}$ be strictly stationary, continuous in probability, and ergodic. Then, there is a process $\{Y(t,\omega^*), -\infty < t < \infty\}$ which is separable and measurable with respect to both t and ω^* such that

$$\mathbf{P}^*\{X(t) = Y(t)\} = 1, -\infty < t < \infty. \tag{10.5.1}$$

Let $\{Y_N(t,\omega^*), -\infty < t < \infty\}$ be defined by

$$Y_N(t,\omega^*) = \begin{cases} Y(t,\omega^*), & |Y(t,\omega^*)| \leqslant N, \\ 0 & \text{otherwise.} \end{cases} \tag{10.5.2}$$

Then Y_N strictly stationarity, continuous in the mean and ergodic. Morever, $Y_N \rightarrow Y$ in the V-topology by the strict stationary of Y as $N \rightarrow \infty$. Note Y_N is *uniformly bounded.*

Next, let

$$Z_{M,N}(t,\omega^*) = \frac{M}{2} \int_{-\frac{1}{M}}^{+\frac{1}{M}} Y_N(t+s,\omega^*)ds. \tag{10.5.3}$$

$Z_{M,N}$ converges to Y_N in the V-topology as $M \rightarrow \infty$, is also strictly stationary continuous in the mean, and is ergodic. Furthermore, $Z_{M,N}$ is *uniformly continuous*:

$$\begin{aligned} &|Z_{M,N}(t+\delta,\omega^*) - Z_{M,N}(t,\omega^*)| \\ &\leqslant \frac{M}{2} \left| \int_{-\frac{1}{M}+t+\delta}^{\frac{1}{M}+t+\delta} Y_N(s)ds - \int_{-\frac{1}{M}+t}^{\frac{1}{M}+t} Y_N(s)ds \right| \\ &\leqslant MN\delta. \end{aligned} \tag{10.5.4}$$

We denote $Z_{M,N}$ the bounded, smoothed version of X, simply as Z.

We now form the *Wiener-Nisio functional* from a typical path $Z(\cdot,\omega_0^*)$ of Z as follows:

$$F_k(t,\omega) = Z(-\lambda_k(\omega_t^+),\omega_0{}^*). \tag{10.5.5}$$

Note that F_k is measurable with respect to (t,ω), where ω corresponds to the space of Wiener increments $dW(t)$. The various signs in (10.5.5) can be reconciled by noting that, as t increases, the Wiener increments process W moves "to the left" by t units, and thus $\lambda_k(\omega_t^+)$ decrements by t units; negating $\lambda_k(\omega_t^+)$ assures that the path Z $(\cdot,\omega^*,)$ is scanned in the forward (increasing time) direction as t increases. (See Figure (10.5.1).)

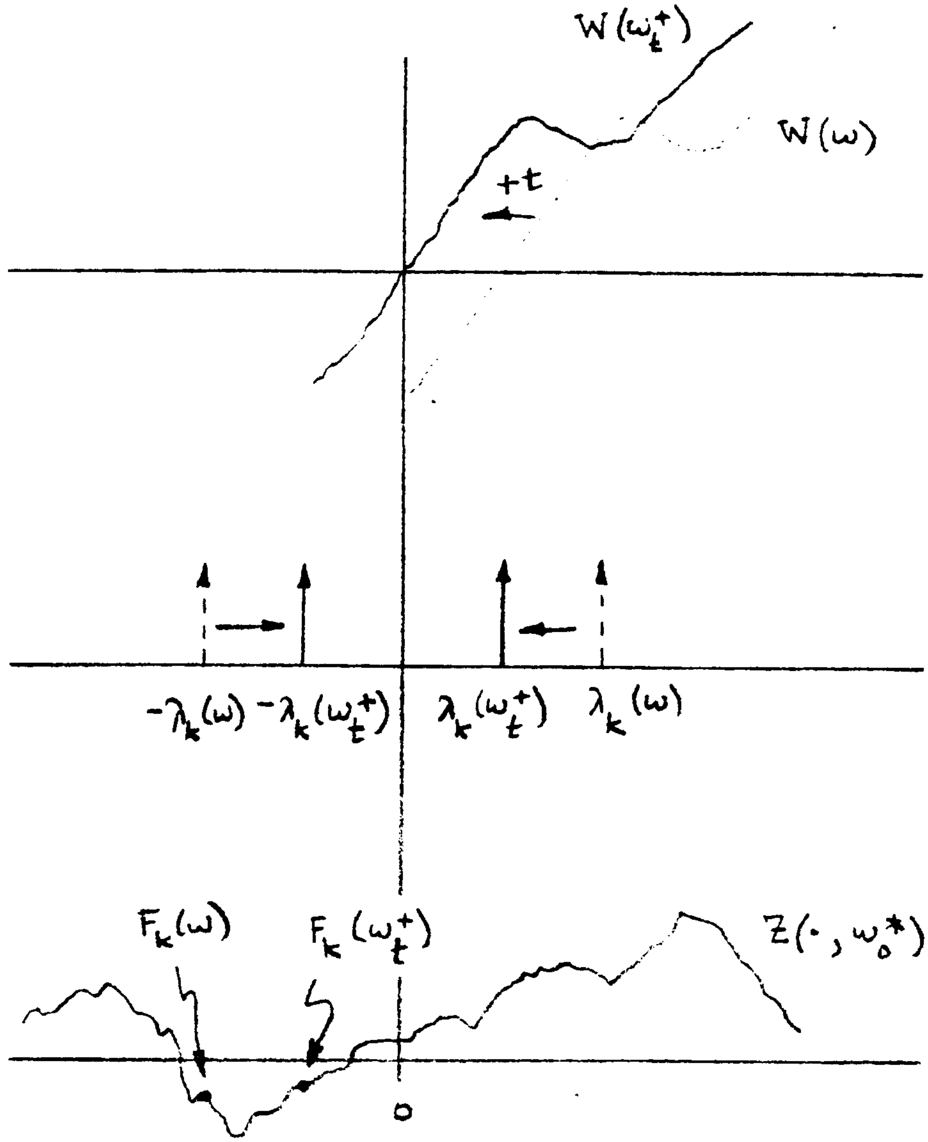


Fig. 10.5.1 — Illustration of sign conventions for shifts $+t$

10.6 Convergence of F_k and Wiener Series in the U-Topology.

To demonstrate convergence of F_k in the U-topology, one computes the characteristic function of $\left\{F_k(t_1,\omega),\ldots, F_k(t_n,\omega)\right\}$:

$$\mathbf{E}\left\{e^{j\,\Sigma\theta_i F_k(t_i,\omega)}\right\} = \mathbf{E}\left\{e^{j\,\Sigma\theta_i\, Z(-\lambda_K(\omega_{t_i}^{+})),\,\omega_0^{*})}\right\}$$

$$= \mathbf{E}\left\{e^{j\,\Sigma\,\theta_i\, Z(-\lambda_K(\omega)+t_i),\,\omega_0^{*})}\right\} + \epsilon, \qquad (10.6.1)$$

the approximations justified by theorem 10.3. Using theorem 10.2, the uniform boundedness and uniform continuity of Z, and the stationary, ergodic properties of Z, one obtains the result that

$$\left| \mathrm{E}\, e^{j\,\Sigma\,\theta_i F_k(t_i,\omega)} - \mathrm{E}\, e^{j\,\Sigma\,\theta_i\, Z(t_i,\omega^{*})}\right| < \epsilon . \qquad (10.6.2)$$

This holds provided that k is greater than both the time extent needed to approximate the characteristic function of Z by time averages over $Z(\cdot,\omega_0^{*})$ and to justify preservation of shifts in theorem 10.3, and if n, $|\theta_i|$, and $t \leqslant \frac{1}{\epsilon}$ (For details, see [NIS1]). This then proves:

$$\left\{F_k(t,\omega),\, -\infty < t < \infty\right\} \rightarrow \left\{Z(t,\omega^{*}),\, -\infty < t < \infty\right\} \qquad (10.6.3)$$

in the sense of the U-topology as $k \rightarrow \infty$,

Next, we bring in Wiener Series. Note that $F_k(0,\omega)$ is an L_2 functional of Wiener increments dW, and therefore admits a Wiener series expansion $G_{k,m}(0,\omega)$ such that:

$$\mathrm{E}\left\{\left|F_k(0,\omega) - G_{k,m}(0,\omega)\right|^2\right\} < \epsilon \text{ for large } m . \qquad (10.6.4)$$

By the stationarity of $dW(t)$, and defining

$$G_{k,m}(t,\omega) = G_{k,m}(0,\omega_t^+), \tag{10.6.5}$$

we see that

$$\mathbf{E}\left\{ \left| F_k(t,\omega) - G_{k,m}(t,\omega) \right|^2 \right\}$$
$$= \mathbf{E}\left\{ \left| F_k(0,\omega_t^+) - G_{k,m}(0,\omega_t^+) \right|^2 \right\} < \epsilon \tag{10.6.6}$$

Thus, $\left\{ G_{k,m}(t,\omega) \right\} \to \left\{ F_k(t,\omega) \right\}$ in the V-topology, and therefore in the U-topology.

We have thus proven that, if process X satisfies (10.0.1), then there exists a family of finite-order Wiener Series which, when driven by the flow of white noise, generate processes which converge to X in the U-topology.

10.7 Relationship of Convergence in U-Topology to Weak Convergence

Since we are dealing with stochastic processes with continuous sample paths, it is natural to investigate whether or note convergence in the U-topology implies convergence in the probability measure on the space of continuous sample paths with the sup norm as the the metric, that is, convergence in the weak sense. As Billingsley [BIL1, BIL2], Kusher [KUS1], and others explain, weak convergence of measures on C, the space of continuous functions, requires more than convergence of the finite distributions. To assure no set mass escapes the scrutiny of finite distribution samplings, something more about the process, such as the notion of tightness, must be established.

We first note that convergence in the U-topology implies convergence of the finite order characteristic functions for each chosen set of parameters θ_i, $i = 1, \ldots, n$.

By [BIL1, theorem 7.6], this is necessary and sufficient for convergence of the finite distributions. We therefore have that convergence in the Nisio U-topology implies convergence in finite-order distributions.

It should be noted that Nisio deliberately favors the stronger notion of the U-topology rather than convergence of finite distributions because the U-topology convergence is transitive ($X \rightarrow Y$, $Y \rightarrow Z \Rightarrow X \rightarrow Z$, while the other (Helly's sense) is not. (See the introduction to [NIS1].)

Now, *sufficient* conditions for a process to be tight are:

- uniform boundedness
- uniform continuity

In constructing the Wiener-Nisio functionals, we relied on these properties of $Z = Z_{M,N}$ in (10.5.3), the smoothed, bounded version of X. We therefore may conclude the following: If a sequence of Wiener-Nisio functionals, and corresponding Wiener series, are developed whose output processes converge to Z in the U-topology, these processes converge weakly to Z. While the output processes converge to X in the U-topology, they do not necessarily converge weakly to X.

The crucial issue is how X behaves relative to Z, its smoothed, bounded version. If we adopt a modeling framework where our only notion of X is based on a simple path $X(\cdot, \omega_0{}^*)$, and this observation path is both time and band-limited, then we are simply incapable of truly determining, on the basis of real data, whether or not X is tight. Simply put, in a practical situation, convergence of finite distributions is all that can be, or need be, demonstrated.

10.8 Causal Modification to the Wiener-Nisio Functional

I now demonstrate how, by slightly altering the construction of the Wiener-Nisio functional, the kernels resulting from the identification procedures of Section 9 will all be one-sided:

$$h_p(t_1, \ldots, t_p) = 0 \text{ for } t_1 > 0, \ldots, t_p > 0, \quad p = 1, 2, \ldots \tag{10.8.1}$$

To determine h_p for a particular Wiener-Nisio functional we will do the following, as summarized in Section 11:

- Pick an ω.
- Generate $dW(t,\omega)$. (10.8.2)
- Compute $\lambda_n(\omega)$ from dW.
- Compute $F_K(\omega)$ from λ_n and $X(\cdot,\omega_0^*)$.
- Compute $F_K(\omega) \cdot Q_{p,\epsilon}(\tau_1, \ldots, \tau_n)(\omega)$for this ω.
- Accumulate over repetitions of $\omega$to approximate (9.1.4).

In the context of procedure (10.8.2), for each ω, the *present* is always $\lambda_n(\omega)$. Therefore, the key to achieving causality in (10.8.1) by this procedure is to assume that $\lambda_n(\omega)$ (and therefore $F_K(\omega)$) depends only on $\{dW(s),\ s \leqslant \lambda_n\}$.

Although the infimum in (10.1.11) might give the impression that $\lambda_n(\omega)$ is future-dependent, this is not so. For each ω, $\lambda_n(\omega)$ is the *first time* that condition T_n in (10.1.2) can be ascertained, and this, in turn, can be determined from $|W(t+1) - W(t-1)|,\ -n \leqslant t \leqslant \lambda_n$. By using

$$S'(\omega) = \{t : |W(t) - W(t-2|\} > 1 \tag{10.8.3}$$

instead of (10.1.1), and extending the working range on $X(\cdot,\omega_0^*)$ by 1-unit, we then achieve the desired goal.

We are the boys/That fear no noise . . .
. . . Oliver Goldsmith

11. An Algorithm for Building Dynamical Stochastic Process Models from Data

I now draw together the preceding ideas into a prescription for building a sequence of stochastic models whose outputs converge in finite distributions to a given process X. The process X must satisfy:

a. $X(t, \omega): R \times \Omega \rightarrow R$ is scalar and real-valued,

b. X is strictly stationary,

c. X is ergodic,

d. X is continuous in probability,

e. X is lossy (see Section 5.4).

The raw data driving the algorithm is a typical sample path of X, denoted $X(\cdot,\omega_o^+)$, of arbitrarily great time extent. It is also assumed unlimited computational power is available, and that the model synthesis can be performed off line (that is, the *realization procedure* need not run in real time). The resulting models, however, are intended to be finite-dimensional *casual* bilinear systems driven by white noise.

Step 1: Process $X(\cdot,\omega_o^+)$ according to (10.5.2) and (10.5.3) to produce a bounded, uniformly continuous path $Z(\cdot,\omega_o^+)$. In any practical situation involving time and band limitation. $X(\cdot,\omega_o^+)$ will automatically qualify as Z without any further filtering. The bound N in (10.5.2) should be chosen to preserve tail probabilities, while the width of the smoothing window M in (10.5.3) should be governed by the projected bandwidth of the application for which the process model is intended.

Step 2: (Synthesis of Wiener Noise functional for each ω). Create a white noise $n(t, \omega)$ running from time t_o to time $t_o + 2 + \lambda_n(\omega)$ as determined below. For each n, generate a corresponding standard Wiener increments process according to

$$W(t,\omega) = \int_{t_0}^{t} n(\xi,\omega)\,d\xi \,.$$

As a running function of time, observe

$$D(t,\omega) = |W(t) - W(t-2)|$$

beginning at $t = t_o + 2$.

Choose n sufficiently large to satisfy Theorems 10.2 and 10.3 in the following manner. Let l in (10.3.1) be chosen much larger than the *memory* extent of X (recall X is lossy). Choose n large enough to guarantee (10.3.1) and also to insure λ is uniformly distributed over many l-intervals of X sufficient to guarantee good approximation of first order statistics of X by uniform, random sampling over $0 \leqslant t \leqslant n$. The former requirement on n, depending on l, can be relaxed somewhat by throwing out all $\lambda_n \leqslant l$ (see 10.3.5), naturally as a tradeoff of required ω repetitions.

Arbitrarily establish a time reference for $n(t)$ and $W(t)$ such that $t_o = -n$. (This is accomplished by a finite shift of time scale.)

Drive a timer initially set to zero as follows: Whenever $D(t,\omega) > 1$, run the timer, while if $D(t,\omega) \leqslant 1$, reset the timer to zero. Note the first instance t when the timer exceeds n. Call this time $\lambda_n(\omega)$. Note that while noise $n(t)$ and Wiener process $W(t)$ can be generated a little ahead of the running time, and stopped as soon as $\lambda_n(\omega)$ is determined.

Set output $Y(n,\omega) = X(-\lambda_n(\omega),\omega_o^*)$.

Step 3a: (Computation of Wiener Kernels). While driving the above apparatus, drive $n(t)$ into smoothing filters $q_\epsilon(t_i)$, $i = 1,\ldots,p$, where p is the order of the Wiener expansion. (The order p should be increased until the derived $p-n$th or kernels are extremely small). The window with ϵ is again based on consideration of the bandwidth of the intended application for the model. For each ω, accumulate values of $Y \cdot Q_{p,\epsilon}(\tau_1,\ldots,\tau_p)$ for values of τ_i that are spaced multiples of ϵ and such that $\tau_i < \tau_j < 0$, $i < j$. Finally, approximate $h_{p,\epsilon}$ by running Steps 2 and 3a over many ω. Note that the "causal" properties of the Wiener-Nisio functional allows us to set h_p=0 for positive τ.

Step 3b: (Alternative - Fourier-Hermite Realization)

Choose a bank of Laguerre filters (see Section 5.5) with a taper appropriate to the process memory, and drive the filters, from 0 initial condition, by $n(t)$. Form Fourier-Hermite functions (5.3.3), and compute the Fourier-Hermite coefficients (5.3.5) by taking products and approximating the correlations by averaging over repetitive ω.

Step 4: (Realize Wiener Series by Bilinear Stochastic system)

Regard each h_p as the kernel (symmetric) of a Volterra system. Use a differentially separable L_2 approximation for each h_p (or use the Fourier-Hermite equivalent.) For each p, realize this Volterra system as a deterministic bilinear system (see Section 8.) Convert each p-th order system to a Wiener G-functional either by computing the *derived* lower Volterra systems and realizing them as well, or use the method described in Section 8 to uncorrect the Volterra system driven by white noise to produce, as an output, the solution to the bilinear system viewed as an Ito differential equation.

"They liked the book the better the more it made them cry." ... Oliver Goldsmith

12. On Realizing Gaussian Processes

This and the following section describe properties of Wiener Series process models for two important classes of processes: Gaussian Processes, which traditionally are viewed as first-order, linear functionals of white noise, and Rayleigh processes, which have second-order or quadratic functional realizations. The point these sections make is that there are significant weaknesses to the procedure outlined in Section 11.

While it is widely known that any stationary Gaussian process which has no trivial past can be realized by a first order linear system driven by white noise, it is not so well known that all white noise-driven causal Wiener Series realizations of Gaussian processes need not be linear. McKean [MCK1] alludes to this, and Nisio [NIS2] provides a detailed description of such series which is reviewed here. I will demonstrate that the procedure in Section 11, based on Wiener-Nisio functionals, when applied to Gaussian processes, will likely result in an infinite-order Wiener series realization. This, I believe, is a serious weakness of the procedure if one hopes to converge rapidly toward *minimal* order realizations particularly in cases where such solutions are known to exist.

12.1 The Canonical Properties of Kernels for Causal Wiener Series Realizations

Let $dW_t = dW(t,\omega)$ be a Wiener process, and Let $B_t(dW)$ denotes the field generated by $\{ W(u) - W(v),\ u,\ v \leqslant t\}$.

For any set $E \in B_0(dW)$, let

$$\chi_E(\omega) = \begin{cases} 1 \text{ on } E \\ -1 \text{ on } E^c. \end{cases} \tag{12.1.1}$$

The indicator χ in effect tags Wiener process paths much like the window indicators in section 5.3. This provides a vehicle for identifying Wiener paths via the trajectories. χ is a functional of increments $dW(t)$, $t \leqslant 0$, and admits a Wiener Series expansion:

$$\chi_E(\omega) = \sum_{n=0}^{\infty} \int_{-\infty}^{0} \cdots \int_{\infty}^{0} \rho_{E,n}(t_1, \ldots, t_n)\, dW_{t_1} \ldots dW_{t_n}, \tag{12.1.2}$$

where $\{\rho_{E,n}\}$ are the kernels associated with $\chi_E(\omega)$.

$$\text{Let } \mathbf{F} = \{\rho_{E,i}\ i = 0,1, \ldots \mid E \in B_0(dW)\}. \tag{12.1.3}$$

F, therefore, is the collection of all Wiener kernel images associated with the indicators χ_E of Wiener increment process trajectories.

Let ρ_i, $i = 0, 1, \ldots, \in \mathbf{F}$, and let

$$\begin{aligned} d\overline{W} &= \overline{W}(t,\omega) - \overline{W}(s,\omega) \\ &= \sum_{n=o}^{\infty} \int_s^t dW_\tau \left[\int_{-\infty}^{\tau} \cdots \int_{-\infty}^{\tau} \rho_n(t_1 - \tau, \ldots, t_n - \tau)\, dW_{\tau_1} \ldots dW_{\tau_n} \right] \end{aligned} \tag{12.1.4}$$

Theorem: $\overline{W}$ is a Wiener increment process.

Proof: See [NIS2, pg. 132]. Note that in 12.1.4,

$$[\ldots] = \chi_E(\omega), \tag{12.1.5}$$

so $\overline{W}$ is derived from W solely by altering the *signs* of the accumulated increments. χ_E is therefore a vehicle for substituting one Wiener increments process for another by means of change in sign of the increments.

We denote the relation

$$\mathbf{a} = \mathbf{b} \circ \rho \tag{12.1.6}$$

to mean

$$\sum_{n=0}^{\infty} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} a_n \ (t_1, \ldots, t_n) \ dW_{t_1} \ldots dW_{t_n}$$

$$= \sum_{n=0}^{\infty} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} b_n \ (t_1, \ldots, b_n) \ d\overline{W}_{t_1} \ldots d\overline{W}_{t_n}, \tag{12.1.7}$$

that is, b is the kernel image of a Wiener series based on $d\overline{W}$, when a is the kernel image of a series based on dW, and dW and $d\overline{W}$ are related by (12.1.4), that is, the paths are related modulo the signs of Wiener increments.

We call a Wiener series realization of process Y with kernels g properly canonical if and only if the span of Y at time t precisely matches the span of dW_t, for all t.

Theorem: [NIS2, pg. 134]. All causal Wiener series realizations of Y have kernels of the form

$$\mathbf{f} = \mathbf{g} \ o \ \rho \ . \tag{12.1.8}$$

Thus, the members of the equivalence class of causal realizations can be traced to the choice of signs for the respective Wiener increments.

12.2 The Order of Wiener Series for Gaussian Processes

Let

$$Y(t,\omega) = \int_{-\infty}^{t} g(t-s) dW_s \tag{12.2.1}$$

be a *Gaussian* process with the proper, canonical realization given in (12.2.1) (i.e., dW_t is the innovations of $Y(t)$).

Theorem: [NIS2, pg. 141] The kernels of all the causal Wiener series realizations of Y are of the form:

$$f_n(t_1, \ldots, t_n)$$
$$= \frac{1}{n} \sum_{i=1}^{n} g(t_i) \chi_{(-\infty, t_i]^{n-1}}(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n) \times$$
$$\rho_{n-1}(t_1 - t_i, \ldots, t_{i-1} - t_i, t_{i+1} - t_i, \ldots, t_n - t_i). \tag{12.2.2}$$

Theorem: The degree of a causal Wiener Series representation of a Gaussian process is either 1 or ∞.

Proof: [NIS2, pg. 141] From (12.2.2),

$$||f_n||^2 = \frac{1}{n} ||g||^2 ||\phi_{n-1}||^2. \tag{12.2.3}$$

It is therefore enough to show that if

$$\rho_E = \{\rho_{E,0}, \rho_{E,1}, \ldots, \rho_{E,K}, 0, 0 \ldots\} \tag{12.2.4}$$

for finite K, then $K = 0$. By the definition of ρ_E,

$$\chi_E(\omega) = \sum_{n=0}^{K} \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \phi_{E,n}(t_1, \ldots, t_n) dW_{t_1} \ldots dW_{t_n}. \tag{12.2.5}$$

Writing χ_E via Fourier-Hermite series based on Gaussian functionals $\xi_n(\omega)$ of dW_t:

$$\xi_n(\omega) = \int_{-\infty}^{0} \alpha_n(t) dW_t, \quad n = 1, 2, \ldots, \tag{12.2.6}$$

we have

$$\chi_E(\omega) = \sum_{p=0}^{K} \sum_{n} \sum_{p_1 + \cdots + p_n = p} \sum_{u_1 \cdots u_n} a_{p_1 \cdots p_n}^{u_1 \cdots u_n} \prod_{i=1}^{n} H_{p_i}\left(\frac{\xi_i(\omega)}{\sqrt{2}}\right). \tag{12.2.7}$$

Expand $\chi_E(\omega)$ as a power series of $\xi_1(\omega)$:

$$\chi_E(\omega) = \sum_{i=0}^{K} c_i(\xi_1, \ldots,) \xi_1^i(\omega). \tag{12.2.8}$$

It is a polynomial of $\xi_1(\omega)$ of degree K at most. But χ is 1 or -1, so the degree of this polynomial must be zero.

We repeat this, and by induction we reach the conclusion that $\chi_E(\omega)$ is a *constant* when K is finite. By (12.2.5), this means K is zero, as we intended to show.

12.3 Order of Realizations Obtained in Section 11 for Gaussian Processes

Theorem: The order = ∞.

Proof: Note that the Wiener-Nisio functional depends on dW only through the *magnitudes* of increments $|W(t) - W(t-2)|$. The corresponding Wiener Series must, therefore, be even order. By the proceeding theorem, this means the order is unbounded.

The significance of this finding is that, even in a situation where we know that a finite-order Wiener expansion model exists for a process (and could perhaps be found by other means), there is little assurance that the procedure in Section 11 will intrinsically converge toward finite-order Wiener expansions.

While truncation of the Wiener series is admissible within the approximation framework, in this particular case one is likely to require an extraordinary number of terms. Recall the Wiener-Nisio functional is even. We are, therefore, compelled to approximate the usual linear functional of $n(t)$ by an ill-suited collection of even order polynomials.

"People who make no noise are dangerous."

Jean de la Fontaine

13. On Realizing Rayleigh Processes

In this section, I investigate further properties of Wiener Series, this time for Rayleigh power processes (the square of Rayleigh processes). There are known second-order Wiener series realizations for these processes. After describing some results due to McKean, I will show that, even if we assume an algorithm like that in Section 11 converges to a finite-order realization, there may be difficulty in resolving critical parameters from data. The parameters are critical in the sense that they radically affect the stability of the realization procedure, but are indeterminate from just analysis of given data.

In what follows, I present McKean's theorem that the maximum order of finite Wiener series can be determined from the tails of the first order density. This theorem agrees with the Nisio theorem (12.2) in that it predicts finite order Gaussian realizations must be linear, and asserts that second-order series give rise to first order densities with exponential tails. I present a second theorem of McKean's which derives the characteristic function of a second-order Wiener series, and from this, shows that second-order Wiener realizations of Raleigh power processes, driven by a single white noise, must have a particular form. I then relate this to the Rayleigh process associated with the sum of squares of *two* independent, identically distributed Gaussian processes. Using this, I show a Wiener series can realize any such Rayleigh process as the envelope of a bandpass Gaussian process. The center frequency of the bandpass filter is shown to be free, and therefore cannot be determined from the Rayleigh process statistics. This implies indeterminacy in the Wiener kernels obtained by the algorithm in Section 11, even for finite series. This is yet another example of how the algorithm in Section 11 is likely to result in models of excessively high order and complexity.

13.1 Theorem: Relation of Max Order to Tails [MCK1]

Let white noise functional Y admit a Wiener series expansion with finite maximum order n. Then

$$e^{-k_1 x^{\frac{2}{n}}} \leqslant \mathbf{P}\left\{ |Y| > x \right\} \leqslant e^{-k_2 x^{\frac{2}{n}}} . \tag{13.1.1}$$

The proof uses the Chernoff bound [GAL] to establish the right hand inequality, and extensions to the work of Eidlin and Linnik [LNN] for the left hand inequality.

13.2 The Characteristic Function of a Second-Order Wiener Series [MCK1].

Let Y be a second order functional of white noise $n(t)$:

$$Y = \Big[\mathbf{E}\{Y\} + \int f_1(t)\, dW_t + \int\int f_2(t_1, t_2)\, dW_{t_1}\, dW_{t_2} , \tag{13.2.1}$$

where the integrals are *Ito* integrals. Regard f_2 as the kernel of a quadratic operator acting on paths, and expand via orthonormal eigenfunctions $\{\mathbf{e}_n\}$ with corresponding eigenvalues $\{\gamma_n\}$:

$$f_2(t_1, t_2) = \sum_{n=1}^{\infty} \gamma_n\, \mathbf{e}_n(t_1)\mathbf{e}_n(t_2) \tag{13.2.2}$$

$$f_1(t) = \sum_{n=1}^{\infty} \beta_n\, \mathbf{e}_n(t) + \mathbf{e}_\infty(t) , \tag{13.2.3}$$

$$\mathbf{e}_\infty \perp \mathbf{e}_n ,\ \ n = 1, 2, \ldots . \tag{13.2.4}$$

Then, letting

$$\xi_n = \int \mathbf{e}_n(t)\, dW_t , \tag{13.2.5}$$

$$\gamma = \sum_{n=1}^{\infty} \beta_n \xi_n + \xi_\infty + \sum_{n=1}^{\infty} \frac{\gamma_n}{2} (\xi_n^2 - 1) , \tag{13.2.6}$$

where the last sum follows because the integrals in (13.2.1) are Ito integrals. Since the $\{\xi_n\}$ are independent and Gaussian, we may write the characteristic function of Y as follows:

$$E\left\{ e^{jkY} \right\} = E\left\{ e^{jk\xi_\infty} \right\} \prod_n E\left\{ e^{jk\,[\beta_n \xi_n + \frac{1}{2}\gamma_n (\xi_n^2 - 1)]} \right\}. \qquad (13.2.7)$$

Now,

$$E\left\{ e^{jk\,[\beta x + \frac{1}{2}\gamma\,(x^2 - 1)]} \right\} =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{jk\,[\beta x + \frac{1}{2}\gamma\,(x^2 - 1)]}\, e^{-\frac{x^2}{2}}\, dx\,. \qquad (13.2.8)$$

But

$$\frac{-x^2}{2} + \frac{j}{2} k\gamma x^2 + jk\,\beta x - \frac{j}{2}\,k\gamma = -x^2 \left(\frac{1 - jk\gamma}{2} \right) + jk\beta x - \frac{j}{2}\,k\gamma \qquad (13.7.9)$$

which is of the form

$$a\,x^2 + b\,x + c \qquad (13.2.10)$$

with

$$a = \left(\frac{1 - jk\gamma}{2} \right) \qquad (13.2.11)$$

$$b = -j\,k\,\beta \qquad (13.2.12)$$

$$c = \frac{-j}{2}\,k\,\gamma\,. \qquad (13.2.13)$$

By completion of squares, (13.2.10) is rewritten as

$$\left(\sqrt{a}\;\, x + \frac{b}{2\sqrt{a}} \right)^2 - \frac{b^2}{4a} + c \qquad (13.2.14)$$

so that (13.2.9) equals:

$$-\left[\left(x\sqrt{\frac{1 - jk\gamma}{2}} - \frac{jk\,\beta}{2\sqrt{\frac{1 - jk\gamma}{2}}} \right)^2 + \frac{k^2\beta^2}{2\,(1 - jk\gamma)} + \frac{jk\gamma}{2} \right]. \qquad (13.2.15)$$

Therefore, (13.2.8) equals:

$$\frac{e^{-\frac{1}{2}\left[\frac{k^2\beta^2}{1-jk\gamma}\right]} e^{-\frac{jk\gamma}{2}}}{(1-jk\gamma)^{\frac{1}{2}}} \tag{13.2.16}$$

as obtained from the formula:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left[\frac{x}{\sqrt{2}\sigma}+C\right]^2} dx = \sigma \tag{13.2.17}$$

with

$$\sigma = \frac{1}{\sqrt{1-jk\gamma}}. \tag{13.2.18}$$

This leads to the result:

$$\mathrm{E}\{e^{jky}\} = e^{\frac{-k^2\alpha^2}{2}} \prod e^{-\frac{1}{2}\left[\frac{k^2\beta_n^2}{1-jk\gamma_n}\right]} \frac{e^{-j\frac{k\gamma_n}{2}}}{(1-jk\gamma_n)^{1/2}}. \tag{13.2.19}$$

13.3 Theorem: Form of Rayleigh Power Finite Series. [MCK1]

If

$$\mathrm{P}\{Y > x\} = \begin{cases} e^{-cx}, & x \geqslant 0, \\ 0, & x < 0, \end{cases} \tag{13.3.1}$$

then the probability density is also one-sided exponential. The Fourier transform has the form:

$$\frac{1}{c+j\omega} \tag{13.3.1}$$

so that the characteristic function has the form:

$$\frac{1}{c-j\omega}. \tag{13.3.2}$$

Examining (13.2.19), we see that there are precisely two orthogonal eigenfunctions $\mathbf{e}_i$, $\mathbf{e}_j$ with identical eigenvalues $\gamma_i = \gamma_j \neq 0$; and that all β's and α must be zero. Further noting that $\mathrm{E}\{Y\} = \frac{1}{c}$, we obtain the result that the second order series for Rayleigh power processes must be of the form:

$$Y = 1 + \int\int g(t_1),g(t_2)dW_{t_1}\, dW_{t_2} + \int\int h(t_1)\; h\;(t_2)\; dW_{t_1}\, dW_{t_2}, \qquad (13.3.3)$$

where

$$g \perp h \qquad (13.3.4)$$

and

$$||g|| = ||h|| = 1. \qquad (13.3.5)$$

Equivalently, we may rewrite (13.3.3) in terms of which noise stochastic integrals:

$$Y = \left[\int g(t)\, n(t)\, dt\right]^2 + \left[\int h(t)\, n(t)\, dt\right]^2, \qquad (13.3.6)$$

where g and h must satisfy (13.3.4) and (13.3.5).

Remarks: First note that Y being of form (13.3.3) or (13.3.6) follows from the *precise* form (Raleigh) of (13.3.1). Specifically, this form does not follow merely from Y having a density with exponential *tail;* note that the function

$$\frac{1}{\sqrt{b-a}}\; e^{-\,at}\, \mathrm{erf}\; (\sqrt{b-a}\,\sqrt{t}\,) \qquad (13.3.7)$$

has an exponential tail, but has a Laplace transform of the form:

$$\frac{1}{(s\, +\, a)(\sqrt{s+b}\,)} \qquad (13.3.8)$$

(see [ABR], pg. 1024), which would imply three distinct eigenfunctions and would not correspond to the form of (13.3.3).

Second, (13.3.6) asserts that if is the sum of the squares of two Gaussian processes which are independent and identically distributed at the same time. There is no constraint on the processes to be independent at differing times:

$$\int g(t_1 - \tau) n(\tau) d\tau \text{ and } \int h(t_2 - \tau) n(\tau) d\tau \qquad (13.3.9)$$

need not be independent when $t_1 \neq t_2$. A particularly noteworthy example is the choice:

$$h = \hat{g}, \qquad (13.3.10)$$

where the tilde denotes the Hilbert transform.

In contrast, in many problems, Y is the square envelope of a *complex* Gaussian process:

$$Y(t) = \left[\int h(t-\tau) n_1(\tau) d\tau\right]^2 + \left[\int h(t-\tau) n_2(\tau) d\tau\right]^2 \qquad (13.3.11)$$

where noises n_1 and n_2 are fully independent, identically distributed. A natural question arises as to how flexible (13.3.3) is at representing any Rayleigh process representation of type (13.3.11). The next section deals with this issue.

13.4 Bandpass Filter Interpretation of Rayleigh Models

I will now review that models of the form (13.3.11) are fully representable as models from class (13.3.10). This follows from analytic signal representations of bandpass processes as complex envelope processes, also known as the Rice representation [HEL].

Recall any stationary Gaussian process can be realized via its spectral representation mapped through the Fourier transform. Using Fig. 13.4.1 as a guide, I define two complex Gaussian processes:

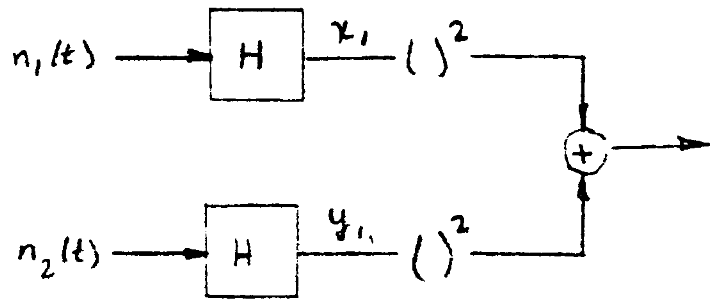
$$z_1(t) = x_1(t) + jy_1(t) \qquad (13.4.1)$$

$$z_2(t) = x_2(t) + jy_2(t) \qquad (13.4.2)$$

where

$$x_1 = n_1 * h \qquad (13.4.3)$$

$$y_1 = n_1 * h \qquad (13.4.4)$$



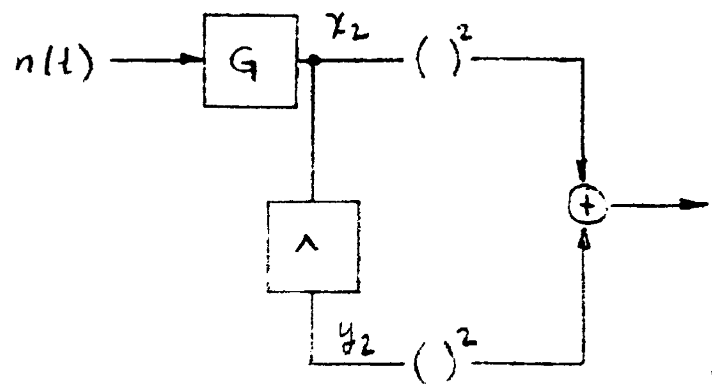


Fig. 13.4.1 — Rayleigh power process models

$$x_2 = n * g \tag{13.4.5}$$

$$y_2 = n * g. \tag{13.4.6}$$

Since $n_1(t)$ and $n_2(t)$ are fully independent, any phase shift in H, the transform of h, can be neglected due to the uniform phase distribution of the pair (n_1, n_2). Therefore, WOLOG, we may assume $H(\omega)$ is purely real. Since h is L_2, $H(\omega)$ can be assumed bandlimited within a controlled, mean-square error. Because x and y, are real, $H(\omega)$ has complex conjugate symmetry.

What I will show is that, if

$$G(\omega) = \frac{H(\omega - \omega_0) + H(-\omega - \omega_0)}{2}, \tag{13.4.7}$$

where ω_0 is larger than the half-bandwidth of H, then z_2 has the same envelope as a complex process z, where z differs from z_1, only by a phase shift. Since

$$\hat{G}(\omega) = -j\left[\frac{H(\omega - \omega_0) - H(-\omega - \omega_0)}{2}\right] \tag{13.4.8}$$

and therefore

$$G(\omega) + j\hat{G}(\omega) = H(\omega - \omega_0), \tag{13.4.9}$$

then z_2 is a complex Gaussian process with single-sided spectrum $|H(\omega - \omega_0)|^2$.

We now form

$$z(t) = z_1(t)e^{-j\omega_0 t}. \tag{13.4.9}$$

The frequency translation in (13.4.9) does not alter the envelope of complex paths of z_2, so z has the same envelope as z_2.

Processes x_2 and y_2 are real which imposes a complex conjugate symmetry constraint on their respective spectral representations relative to zero frequency. However, there is no constraint imposed on the upper versus the lower sidebands of z_2, and therefore, of the sidebands of z. The process z in fact is a complex Gaussian process with a symmetric spectral density, but with independent, identically distributed upper and lower spectral sub-components. This, in turn, implies z can be written as a complex process with independent, identically distributed Gaussian real and imaginary components. Because of (13.4.9), this means that the real and imaginary components of z_2 have the same statistics as those of z, and therefore z_1 and z_2 have the same envelope statistics.

13.5 Impact on Realization Procedure

In what preceded, the choice of center frequency ω_0 was *arbitrary*. The statistics of the envelope of a bandpass process are not dependent on the center frequency of that process. Viewed from a different perspective, the center frequency cannot be determined from the statistics of the bandpass process envelope alone.

What this implies is that the procedure in Section 11 may exhibit erratic behavior in determining second-order kernels even in cases where the adopted series order is low. There is no inherent safeguard to prevent the adoption of an ω_0 of arbitrarily large value. This implies the corresponding second-order kernels could be extraordinarily complex and fine structured, corresponding to very high "Q" realizations.

"He who sleeps in continual noise is wakened by silence." ... William Dean Howells.

14. Concluding Remarks

A careful analysis of criticisms in the last two sections shows the principal weakness to the algorithm of Section 11 is the Wiener-Nisio functional as the chosen candidate realization. Wiener series models per se are really not at fault. I would therefore invite the reader to suggest other candidate functionals of white noise, or auxiliary means to assist the Wiener-Nisio functional, with the goal being reduced model order and complexity.

Recall that the Wiener-Nisio functional does not depend on the sign of the Wiener increments. I believe that one could put more "energy" into lower order and lower band width kernels by assigning a polarity to the functional based on a projection of each white noise path in a particularly chosen direction. For example, one might use the sign of the p-th order integrator driven by $n(t)$ as an appropriate indicator of sign for that particular ω. This approach, I believe, would benefit further from knowledge about the first or higher order densities of the process. For example, this approach might work well for Gaussian processes, which are symmetric, but might require modification for half-sided density processes, the modification being done on a case-by-case basis.

Perhaps the reader will benefit from why I pursued the Wiener-Nisio functional approach in the first place. The driving force was to find some means to make Wiener series work in the general stochastic modelling problem—after all, this appeared to be the original concept and intent behind the Homogeneous Chaos. From considerable ruminating on McKean's article [MCK1], it became increasingly clear that the force behind McKean's approximation claims for Wiener series (that they approximate any stationary process in law) was drawn from Nisio's work [NIS1].

In parallel with the Wiener-Nisio functional approach, a second approach was being attempted, both by me and a colleague of mine, Robert Gover. John Baras pointed out that the moments of finite Wiener series models obey bilinear differential equations. This observation was based on work by Brockett [BRO1, BRO2]. Baras suggested an approach might be to estimate the multivariate moments, determine the differential equations they satisfy by bilinear system identification methods, and then obtain a set of Wiener kernels from the system of the moment differential equations. A first step was to determine the minimum number of moments needed to synthesize a Wiener series of given order. This entire appraoch might have progressed further had it not been for the discovery and comprehension of Grunbaum's theorem [GRU]: To reconstruct a Gaussian process from its square (representable as an elementary homogeneous second order Wiener series), one would need *all* moments of the square. This seemed to imply that moments of a process might not be the most efficient statistic. At that point, the Wiener-Nisio functional approach, using a typical path of the actual data in a direct manner, seemed more inviting.

I hope all of this conveys to the reader my conviction that there are many fruitful ideas, some merely set aside, some yet undiscovered, that could be pursued with benefit in non-Gaussian process modelling.

Bibliography:

ABR. Milton Abramowitz and Irene A. Segun, editors, *Handbook of Mathematical Functions*, New York: Dover Press, 1968.

ARE. Richard Arens, "Complex Processes for Envelopes of Normal Noise," *IRE Transactions on Information Theory*, September, 1957. pp. 204-207.

AST. Karl J. Astrom, *Introduction to Stochastic Control Theory*, New York: Academic Press, 1970.

BAH. R.S. Baheti, R.R. Mohler, and H.A. Spand III," A New Cross Correlation Algorithm for Volterra Kernel Estimation of Bilinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 4, August 1979. pp. 661-664.

BAR1. J.F. Barrett, "Hermite Functional Expansions and the Calculation of Output Autocorrelation and Spectrum for any Time-invariant Non-linear System with Noise Input," *Journal of Electronics and Control*, Vol. 16, 1964. pp. 107-113.

BAR2. J.F. Barrett, "The Use of Functionals in the Analysis of Nonlinear Physical Systems," *Journal of Electronics and Control*, Vol. 15, 1963. pp. 567-615.

BED. Edward Bedrosian and Stephan O. Rice, "The Output Properties of Volterra Systems (Nonlinear Systems with Memory) driven by Harmonic and Gaussian Inputs," *Proceedings of the IEEE*, Vol. 59, No. 12, December 1971. pp. 1688-1707.

BIL1. Patrick Billingsley, *Convergence of Probability Measures*, John Wiley and Sons, New York, 1968.

BIL2. Patrick Billingsley, *Weak Convergence of Measures: Applications in Probability* Philadelphia: SIAM Monographs, 1971.

BOR. A.A. Borovkov, "Theorems on the Convergence to Markov Diffusion Processes," *Z. Wahrscheinlichkeitstheorie verw. Geb*, 16, 1970. pp. 47-76.

BOY. Stephen Boyd, L.O. Chua, and C.A. Desoer, "Analytic Foundations of Volterra Series," *IMA Journal of Mathematical Control and Information*, Vol. 1, 1984. pp. 243-282.

BRI. David R. Brillinger, "The identification of a particular nonlinear time series system," *Biometrika* Vol. 64, No. 3, 1977. pp. 509-515.

BRO1. Roger W. Brockett, "Nonlinear Systems and Differential Geometry," *Proceedings of the IEEE*, Vol. 64 No. 1, 1976. pp. 61-72.

BRO2. Roger W. Brockett, "Volterra Series and Geometric Control Theory," *Automatica,* Vol 12., 1976. pp. 167-176.

BRO3. Roger W. Brockett and Elmer G. Gilbert, "An Addendum to Volterra Series and Geometric Control Theory," *Automatica*, Vol 12, 1976. p. 635.

BUS. Julian J. Busgang, Leonard Ehrman, and James W. Graham, "Analysis of Nonlinear Systems with Multiple Inputs," *Proceedings of the IEEE*, Vol. 62, No. 8, August 1974. pp. 1088-1119.

CAM. R.H. Cameron and W.T. Martin, "The Orthogonal Development of Non-linear Functionals in Series of Fourier-Hermite Functionals," *Annals of Mathematics,* Vol. 48, No. 2, April 1947. pp. 385-392.

CAR. Gianfranco L. Cariolaro and Giovanni B. DiMasi, "Second-Order Analysis of the Output of a Discrete-Time Volterra System Driven by White Noise," *IEEE Transactions on Information Theory*, Vol. IT-26, No. 2, March 1980. pp. 175-184.

CHA. Daniel Chambers and Eric Slud, "Central Limit Theorems for Nonlinear Functionals of Stationary Gaussian Processes," TR 85-3, MD 85-4-DC & ES, University of Maryland Mathematics Departments, January 18, 1985.

CLA1. Steven J. Clancy and Wilson J. Rugh, "On the Realization Problem for Stationary, Homogeneous, Discrete-Time Systems," John Hopkins University.

CLA2. Steven J. Clancy and Wilson J. Rugh, "A Note on the Identification of Discrete-Time Polynomial Systems," *IEEE Transactions on Automatic Control,* Vol. AC-24, No. 6, December 1979. pp. 975-978.

CRO1. P.E. Crouch, *Dynamical Realizations of Finite Volterra Series* (Thesis), Harvard University, Cambridge Mass., August 1977.

CRO2. P.E. Crouch, "Dynamical Realizations of Finite Volterra Series," Control Theory Center Report No. 72, University of Warwick, Coventry CV4 7AL, England .

CHU. Kai Lai Chung, *A Course in Probability Theory*, Second Edition. New York: Academic Press, 1968.

DEF. Rui J.P. de Figueiredo and Thomas A.W. Dwyer, III, "A Best Approximation Framework and Implementation for Simulation of Large-Scale Nonlinear Systems," *IEEE Transactions on Circuits and Systems*, Vol. CAS-27, No. 11, Novemeber 1980. pp. 1005-1080.

DIB. M.D. DiBenedetto and A. Isidori, "Triangular Canonical Forms for Bilinear Systems," *IEEE Transactions on Automatic Control,* Vol. LAC-23, No. 5, October 1978, pp. 877-880.

DOO. J.L. Doob, *Stochastic Processes*, New York: John Wiley and Sons, 1953.

DUG. J. Dugundji, "Envelopes and Pre-Envelopes of Real Waveforms," *IRE Transactions on Information Theory*, March 1958. pp. 53-57.

DWY. T.A.W. Dwyer, III, G.N. Smit, and R.J.P. de Figueiredo, "Input-Output Based Design of Nonlinear Dynamical Systems: Application to the Cancelation of Nonlinear Distortion," Proceedings of the 1981 IEEE International Symposium on Circuits and Systems, Chicago, Illinois, April 27-29, 1981.

DYM. H. Dym and H.P. McKean, *Fourier Series and Integrals*, New York: Academic Press, 1972.

FLI. Michel Fliess, "A Note on Volterra Series Expansions for Nonlinear Differential Systems," *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 1, February 1980. pp. 116-117.

FRI. Avner Friedman, *Stochastic Differential Equations and Applications Vol. I*, New York: Academic Press, 1975.

GAL. Robert G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley and Sons, 1968.

GIL1. Elmer G. Gilbert, "Functional Expansions for the Response of Nonlinear Differential Systems," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 6, December 1977. pp. 909-921.

GIL2. Elmer G. Gilbert, "Bilinear and 2-Power Input-Output Maps: Finite Dimensional Realizations and the Role of Functional Series," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 3, June 1978. pp. 418-425.

GRU. F. Alberto Grunbaum, "The Square of a Gaussian Process," *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 23, 1972. pp. 121-124.

HAR. C.J. Harris, *A Historical Literature Review of Polynomial Operator Theory*, Control Systems Center Report No. 340, July 1976. Control Systems Center, UMIST, Sackville Street, Manchester.

HEL. Carl W. Helstrom, *Statistical Theory of Signal Detection*, Second Edition, New York: Pergamon Press, 1968.

HER. Robert Hermann and Arthur J. Krener, "Nonlinear Controllability and Observability," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 5, October 1977. pp. 728-740.

HID1. Takeyuki Hida, "Analysis of Brownian Functionals," Carleton Mathematical Lecture Notes No. 13, Second Edition, 1978.

HID2. Takeyuki Hida, "Complex White Noise and Infinite Dimensional Unitary Group," Mathematical Institute, Nagoya University Lecture Notes, 1971.

HID3. Takeyuki Hida, "Causal Analysis in Terms of Brownian Motion," *Multivariate Analysis — V*, P.R. Krishnaiah, ed., North-Holland Publishing Company, 1980. pp. 111-118.

HID4. Takeyuki Hida, "Generalized Multiple Wiener Integrals," *Proceedings of the Japan Academy*, Vol. 54, Ser. A, No. 3, 1978. pp. 55-58.

HID5 Takeyuki Hida, "Nonlinear Brownian Functionals," Proceedings of the 18-th IEEE Conference on Decision and Control, Dec. 12-14, 1979. pp. 326-328.

HUA1. Steel T. Huang and Stamatis Cambanis, "Stochastic and Multiple Wiener Integrals for Gaussian Processes," *The Annals of Probability*, Vol. 6, No. 4, 1978. pp. 585-614.

HUA2. Steel T. Huang and Stamatis Cambanis, "On the Representation of Nonlinear Systems with Gaussian Inputs," *Stochastics*, Vol. 2, 1979. pp. 173-189.

HUA3. Steel T. Huang and Stamatis Cambanis, "Spherically Invariant Processes: Their Nonlinear Structure, Discrimination, and Estimation," *Journal of Multivariate Analysis*, Vol. 9, No. 1, March 1979, pp. 59-83.

ISI1. Alberto Isidori, "Direct Construction of Minimal Bilinear Realizations from Nonlinear Input-Output Maps," *IEEE Transactions on Automatic Control*, Vol. AC-18, No. 6, December 1973. pp. 626-631.

ITO1. Kiyosi Ito, "Multiple Wiener Integral," *Journal of the Mathematical Society of Japan*, Vol. 3, No. 1, May 1951. pp. 157-169.

ITO2. Kiyosi Ito and Henry P. McKean, Jr., *Diffusion Processes and Their Sample Paths*, New York: Academic Press, 1965.

JIM E. Jimenez, "Comments on Triangular Canonical Forms for Bilinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 1, February 1980. pp. 137-138.

KAC. Mark Kac and A.J.F. Siegert, "On the Theory of Noise in Radio Receivers with Square Law Detectors," *Journal of Applied Physics* Vol. 18, April, 1947. pp. 383-397.

KAS. R.L. Kashyap and A. Ramachandra Rao, *Dynamic Stochastic Models from Empirical Data*, New York: Academic Press, 1976.

KUS1. Harold J. Kushner, *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, 1977. Chapter 2.

KUS2. Harold J. Kushner, *Introduction to Stochastic Control*, New York: Holt, Rinehart and Winston, 1971.

LEE. Y.W. Lee, *Statistical Theory of Communication* New York: John Wiley and Sons, 1960.

LES. Casimir Lesiak and Arthur J. Krener, "The Existence and Uniqueness of Volterra Series for Nonlinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 6, December 1978. pp. 1090-1095.

LIG. Thomas M. Liggett, "On Convergent Diffusions: The Densities and the Conditioned Processes," *Indiana University Mathematics Journal*, Vol. 20, No. 3, 1970. pp. 265-279.

LIN. Anders Linquist, Sanjoy Mitter, and Giorgio Picci, "Toward a Theory of Nonlinear Stochastic Realization," Proceedings of the Workshop in Feedback and Synthesis of Linear and Nonlinear Systems, Bielefeld, West Germany, June 22-26, 1981, and Rome, Italy, Jun 29-July 3, 1981. Springer-Verlag.

LNN. Yu V. Linnik and V.L. Eidlin, "Remark on Analytic Transfomations of Normal Vectors," *Theory of Probability and its Applications* (SIAM), Volume XIII, No. 3, 1968. pp. 707-710.

LIP. R.S. Liptser and A.N. Shiryayev, *Statistics of Random Processes,* Part I: General Theory, New York: Springer-Verlag 1977.

MAR1. Steven I. Marcus and Alan S. Willsky, "Algebraic Structure and Finite Dimensional Nonlinear Estimation," *SIAM J. Math. Anal.*, Vol. 9, No. 2, April 1978. pp. 312-327.

MAR2. Steven I. Marcus, Sanjoy K. Mitter, Daniel Ocone, "Finite Dimensional Nonlinear Estimation in Continuous and Discrete Time," International Conference on Analysis and Optimzation of Stochastic Systems, Oxford, September, 1978.

MRM. Panos Z. Marmarelis and Ken-Ichi Naka, "White-Noise Analysis of a Neuron Chain: An Application of the Wiener Theory," *Science* Vol. 175, 17 March 1972. pp. 1276-1278.

MAS1. Elias Masry and Stamatis Cambanis, "Signal Identification after Noisy Nonlinear Transformations," *IEEE Transactions on Information Theory*, Vol. IT-26, No. 1, January 1980. pp. 50-58.

MAS2. Elias Masry and Stamatis Cambanis, "On the Reconstruction of the Covariance of Stationary Gaussian Processes Observed Through Zero-Memory Nonlinearities," *IEEE Transactions on Information Theory*, Vol. IT-24, No. 4, 1978. pp. 485-494.

MAS3. Elias Masry and Stamatis Cambanis, "On the Reconstruction of the Covariance of Stationary Gaussian Processes Observed Through Zero-Memory Nonlinearities—Part II," *IEEE Transactions on Information, Theory*, Vol. IT-26, No. 4, July 1980. pp. 503-507.

MCK1. Henry P. McKean, Jr., "Wiener's Theory of Nonlinear Noise," SIAM Proceedings on Stochastic Differential Equations, Vol. VI, 1973. pp. 191-209.

MCK2. Henry P. McKean, Jr., "Geometry of Differential Space," *The Annals of Probability*, 1973, Vol. 1, No. pp. 197-206.

MCK3. Henry P. McKean, Jr., *Stochastic Integrals*, New York: Academic Press, 1969.

MCS. E.J. McShane, *Stochastic Calculus and Stochastic Models*, New York: Academic Press, 1974.

MIT1. Sanjoy K. Mitter, "On the Analogy Between Mathematical Problems on Nonlinear Filtering and Quantum Physics," *Ricerche di Automatica*, Vol. 10, No. 2, December 1979. pp. 163-216.

MIT2. S.K. Mitter and D. Ocone, "Multiple Integral Expansions for Nonlinear Filtering," Proceedings of the 18th IEEE Conference on Decision and Control, December 12-14, 1979, Fort Lauderdale, Florida. pp. 329-334.

MIZ1. Glenn E. Mitzel, Steven J. Clancy and Wilson J. Rugh, "On Transfer Function Representations for Homogeneous Nonlinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 2, April 1979. pp. 242-249.

MIZ2. Glenn E. Mitzel and Wilson Rugh, "On a Multidimensional S-Transform and the Realization Problem for Homogeneous Nonlinear Systems," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 5, October 1977. pp. 825-830.

NIS1. Makiko Nisio, "On Polynomial Approximation for Strictly Stationary Processes," *Journal of the Mathematical Society of Japan*, Vol. 12, No. 2, April 1960. pp. 207-226.

NIS2. Makiko Nisio, "Remarks on the Canonical Representation of Strictly Stationary Processes," J. Math., Kyoto University, Vol. 1 No. 1, 1961. pp. 129-146.

PAL. G. Palm and T. Poggio, "The Volterra Representation and the Wiener Expansion: Validity and Pitfalls," *SIAM J. Appl. Math.*, Vol. 33, No. 2, September 1977. pp. 195-216.

POR. William A. Porter, "An Overview of Polynomic System Theory," *Proceeding of the IEEE*, Vol. 64, No. 1, January 1976. pp. 18-23.

REE. I.S. Reed, "On A Moment Theorem for Complex Gaussian Processes," *IRE Transactions on Information Theory*, April 1962. pp. 194-195.

ROO. W.L. Root, "On System Measurement and Identification," Proceedings *Symposium on System Theory*, Polytechnic Institute of Brooklyn, April 20-22, 1965. pp. 133-157.

RUB. Antonio Ruberti, Alberto Isidori, and Paolo D'Alessandro, *Theory of Bilinear Dynamical Systems*, Springer-Verlag, 1972.

RUD. Michael Rudko and Donald D. Wiener, "Volterra Systems with Random Inputs: A Formalized Approach," *IEEE Transactions on Communications*, Vol. COM-26, No. 2, February 1978. pp. 217-227.

RUG1. Wilson J. Rugh, *Nonlinear System Theory — The Volterra/Wiener Approach*, The Johns Hopkins University Press, Baltimore, 1981.

RUG2. Wilson J. Rugh, "On the Construction of Minimal Linear-Analytic Realizations for Homogeneous Systems," John-Hopkins University.

SHE1. Martin Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John-Wiley and Sons, New York, 1980.

SHE2. Martin Schetzen, "Nonlinear System Modeling Based on the Wiener Theory," *Proceedings of the IEEE*, Vol. 69, No. 12, Dec. 1981. pp. 1557-1573.

SHU. Zeev Schuss, *Theory and Applications of Stochastic Differential Equations*, New York: John Wiley and Sons, 1980.

VAN. Harry L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part III: Radar-Sonar Signal Processing, etc., New York: John Wiley and Sons, 1971.

VOL. V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations*, Dover Publications, Inc., New York, 1959.

WAX. Nelson Wax, ed., *Selected Papers on Noise and Stochastic Processes*, New York: Dover Press, 1954.

WIE1. Norbert Wiener, "The Homogenous Chaos," *Norbert Wiener: Collected Works with Commentaries,* Vol. I, P. Masani, ed., The MIT Press, Cambridge, Mass., 1976. pp. 572-611.

WIE2. Norbert Wiener, *Nonlinear Problems in Random Theory*, The Technology Press of MIT and John Wiley and Sons, Inc., New York, 1958.

WON. Eugene Wong, *Stochastic Processes in Information and Dynamical Systems*, New York: McGraw-Hill 1971.

YAS1. Syozo Yasui, "Stochastic Functional Fourier Series, Volterra Series, and Nonlinear Systems Analysis," *IEEE Transactions on Automatic Control*, VO1. AC-24, No. 2, April, 1979. pp. 230-242.

YAS2. Syozo Yasui, "Wiener-Like Fourier Kernels for Nonlinear System Identification and Synthesis," *IEEE Transactions on Automatic Control*, Vol. AC-27, No. 3, June 1982. pp. 677-685.

YOS. Kosaku Yosida, *Functional Analysis*, Sixth Edition, New York: Springer-Verlag, 1980.