A MARTINGALE APPROACH TO RANDOM FIELDS

E. Wong

University of California, Berkeley
and Harvard University

1. INTRODUCTION

A random field $\{X_t, t \in T\}$ is a stochastic process with a multi-dimensional parameter, i.e. $T \subset R^n$. As in the one dimensional case, one is often interested in signal processing problems involving an observation equation of the form

(1.1) $$\eta_t = X_t + \zeta_t , \quad t \in T$$

where η is the observed process, X is the signal, and ζ represents a corrupting noise. In one dimension the noise process is often associated with the superposition of a large number of pulses of comparable amplitudes and of extremely short durations. Such situations (e.g. shot noise) give rise to a process ζ_t which is Gaussian and white. By a white noise we mean a process with a correlation function given by

(1.2) $$E\zeta_t\zeta_s = N_o\delta(t-s)$$

when δ is the Dirac delta function. Of course, Gaussian white noise is an abstraction and to deal with it effectively has required the development of a stochastic calculus, which began in the form of Wiener integrals and acquired its full generalization and power in recent years as a calculus of martingales [1]. The critical step was taken by Ito, who established the connection between "white noise integrals" and martingales, and derived the differentiation rule which is basic to the stochastic calculus.

Since Gaussian white noise is but an abstraction and since to handle it requires a rather elaborate machinery, the question arises as to why it is so widely used. The answer lies in the fact that with appropriate interpretations it can be said that the values of a Gaussian white noise at different times are statistically independent, and this independence gives rise to a major simplification to the analysis of those signal processing problems where the noise is assumed to be white and Gaussian. Results which owe their existence to the simplification include some of the best known formulas of filtering and detection theory.

For processes with a multidimensional parameter, i.e., for random fields, a similar motivation for using Gaussian white noise as a model exists. There is no difficulty in extending the definition of white noise (1.2) to the multidimensional case. The difficulty lies in develiping a stochastic calculus to deal with it. To generalize martingales and their calculus to multi-parameter processes turned out to be far from straightforward. However, for the two-dimensional case, the essential elements of the stochastic calculus are now known, and some preliminary results on applying the calculus to problems in filtering and detection are also in hand. While the full extent of its usefulness remains to be seen, a martingale approach to random fields can now be said to exist, at least for the two-parameter case.

2. WIENER PROCESS AND MARTINGALES

Our first task is to make precise the idea of a white Gaussian noise. In one dimension this is done by viewing it as the formal derivative of a Wiener process or Brownian motion. We shall employ the same device for the multidimensional case. Let R_+^2 denote the positive quadrant of the plane R^2, and let $\{W_t, t \in R_+^2\}$ be a two-parameter Gaussian process with zero mean and a covariance function given by

$$EW_{(t_1,t_2)}W_{(s_1,s_2)} = N_o \min(t_1,s_1)\min(t_2,s_2) \tag{2.1}$$

We shall call N a Wiener process. If we denote by A_t the rectangle to the left and below t then we can view W_t formally as

$$W_t = \int_{A_t} \zeta_s ds \tag{2.2}$$

where ζ_t is a Gaussian white noise, alternatively

$$\zeta_t = \frac{\partial^2}{\partial t_1 \partial t_2} W_t \qquad t = (t_1,t_2) \tag{2.3}$$

It is useful to define a process $\{W(A)\}$ parameterized by Borel sets A in the plane as a Gaussian process with zero mean and covariance property

$$(2.4) \qquad EW(A)W(B) = N_o \text{Area}(A \cap B)$$

If we set $W_t = W(A_t)$ then W_t is a Wiener process. Alternatively, given a Wiener process W_t, a set-parametered process $W(A)$ can be defined in terms of it. Hence, $\{W(A)\}$ and $\{W_t\}$ are equivalent and we shall refer to $W(A)$ also as a Wiener process. Formally, we have

$$(2.5) \qquad W(A) = \int_A \frac{\partial^2 W_t}{\partial t_1 \partial t_2} dt_1 dt_2 = \int_A \zeta_t dt$$

The values of $W(A)$ for nonoverlapping areas are independent and this captures the independence property of white Gaussian noise.

For two points t and s in the plane, denote

$$(2.6) \qquad t \succ s \quad \text{if } t_1 \geq s_1 \text{ and } t_2 \geq s_2$$

and

$$(2.7) \qquad t \curlywedge s \quad \text{if } t_1 \leq s_1 \text{ and } t_2 \geq s_2$$

The relation $\succ$ is a partial ordering for points in R_+^2 with respect to which martingales can now be defined. Let T be a rectangle in R_+^2 of the form $T = \{s: 0 \leq s_1 \leq a,\ 0 \leq s_2 \leq b\}$. We say a random field $\{W_t,\ t \in T\}$ is a martingale if

$$(2.8) \qquad t' \succ t \text{ implies that } E\{M_{t'} | M_s,\ s \prec t\} = M_t \quad \text{w.p.1}$$

We can generalize the concept of a martingale by introducing a family of σ-fields. We shall say that $\{\mathscr{F}_t,\ t \in T\}$ is an increasing family of σ-fields if

$$t' \succ t \text{ implies } \quad \mathscr{F}_{t'} \supset \mathscr{F}_t$$

A random field $\{M_t,\ t \in T\}$ is said to be adapted to $\{\mathscr{F}_t, t \in T\}$ if for each t M_t is $\mathscr{F}_t$-measurable and it is said to be a martingale with respect to $\{\mathscr{F}_t,\ t \in T\}$ of it is adapted and if

$$(2.8') \qquad t' \succ t \text{ implies that } E\{M_{t'} | \mathscr{F}_s,\ s < t\} = M_t \quad \text{w.p.1}$$

A Wiener process $\{W_t,\ t \in R_+^2\}$ is a martingale under definition (2.8). Further if $\{\mathscr{F}_t,\ t \in R_+^2\}$ is an increasing family of σ-fields such that $W(A)$ is $\mathscr{F}_t$-independent whenever A does not intersect A_t, then $\{W_t,\ t \in R_+^2\}$ is a martingale with respect to $\{\mathscr{F}_t,\ t \in R_+^2\}$ and we shall say for such a case that $\{W_t, \mathscr{F}_t,\ t \in R_+^2\}$ is a Wiener process.

3. STOCHASTIC INTEGRATION

Let $\{W_t, \mathscr{F}_t,\ t \in T\}$ be a Wiener process. The first integrals defined with respect to W were of the form

$$(3.1) \qquad \int_T f(s)W(ds)$$

where f is a deterministic function which satisfies

$$(3.2) \qquad \int_T f^2(s)\,ds < \infty$$

Integrals of the form (3.1) are known as Wiener integrals and have long been known.

The first generalization to (3.1) was to replace the deterministic integrand by a random integrand. Let $\{\phi_t, t \in T\}$ be an $\mathscr{F}_t$-adapted random field which satisfies

$$(3.3) \qquad \int_T E\phi_s^2\,ds < \infty$$

Then the integral

$$(3.4) \qquad M = \int_T \phi_s W(ds)$$

can be defined as a straightforward generalization of the Ito integral. The integral M is defined as the quadratic-mean limit of a sequence of approximating sums, i.e.,

$$(3.5) \qquad M = \lim_{n \to \infty} \text{ in q.m. } \sum_{i,j} \phi(t_{ij}^{(n)})W(\Delta_{ij}^{(n)})$$

where $\{t_{ij}^{(n)}\}$ is a rectangular partition of T for each n, $\Delta_{ij}^{(n)}$ denotes the increment

$$(3.6) \qquad \Delta_{ij}^{(n)} = t_{i+1,j+1}^{(n)} - t_{i+1,j}^{(n)} - t_{i,j+1}^{(n)} + t_{ij}^{(n)}$$

and

$$(3.7) \qquad \max_{i,j} \Delta_{ij}^{(n)} \xrightarrow[n \to \infty]{} 0$$

One of the most important properties of the Ito integral preserved in the generalization (3.4) is its martingale property, which can be stated as

$$(3.8) \qquad E(M|\mathscr{F}_t) = \int_{A_t} \phi_s W(ds)$$

Equation (3.5) implies that if we define

$$(3.9) \qquad M_t = \int_{A_t} \phi_s W(ds)$$

then $\{M_t, \mathscr{F}_t, t \in T\}$ is a martingale.

The generalization of the Ito integral to the two-parameter case (indeed to the n-parameter case) is routine. By itself, it is hardly sufficient to yield a usable two-parameter stochastic calculus. For that we need certain completeness results and a differentiation formula. These proved to be far more elusive, and a great deal more of the structure underlying two-parameter martingales had to be uncovered before the desired results were finally derived.

The first completeness results were obtained by Wong and Zakai, who posed and ansered the following question in [1]:

Q_1 Suppose that $\{W_t, t \in T\}$ is a Wiener process, and $\mathscr{F}_{Wt}$ denotes the σ-field generated by $\{W_s, s \in A_t\}$. Let $\{M_t, t \in T\}$ be an $\mathscr{F}_{Wt}$-martingale such that $EM_T^2 < \infty$. Question: is M_t necessarily representable as a stochastic integral of the form (3.9)?

In one dimension the answer is "yes", and this completeness result represents an important fundamental property of the Ito integral. For the two-parameter case the answer turned out to be "no". To represent every $\mathscr{F}_{Wt}$-martingale, we need not only stochastic integrals of the form (3.4) but stochastic integrals of a second type, which we shall write in the form

$$\int_{T\times T} \psi_{s,s'} W(ds)W(ds') \tag{3.10}$$

Observe that it is an integral on T×T, hence a four-fold integral. Clearly, (3.10) needs to be defined, but we shall postpone doing so for a moment. Henceforth, we shall refer to (3.4) as a type-I stochastic integral and (3.10) as a type-II integral. Wong and Zakai showed that every square-integrable $\mathscr{F}_{Wt}$-martingale can be represented in the form

$$M_t = M_0 + \int_{A_t} \psi_s W(ds) + \int_{A_t \times A_t} \psi_{s,s'} W(ds)W(ds') \tag{3.11}$$

where 0 denotes the origin. (3.11) provides a full answer to Q_1.

We now return to the problem of defining (3.10). We shall define a general multiple integral of the form

$$I(\psi;\mu,\nu) = \int_{T\times T} \psi_{s,s'} \mu(ds)\nu(ds') \tag{3.12}$$

where μ and ν can each be either the Lebesgue measure or a Wiener process, $\psi_{s,s'}$ is $\mathscr{F}_{s\vee s'}$-measurable for each (s,s'), and

$$\int_{s\wedge s'} E\psi^2_{s,s'} ds ds' < \infty \tag{3.13}$$

where we recall that $s \wedge s'$ means ($s_1 \leq s_1'$ and $s_2 \geq s_2'$). $I(\psi;\mu,\nu)$ is defined by

$$I(\psi;\mu,\nu) = \lim_{n\to\infty} \text{ in q.m.} \sum_{\substack{i<k\\ j>\ell}} \psi(t_{ij}^{(n)},t_{k\ell}^{(n)})\mu(\Delta_{ij}^{(n)})\nu(\Delta_{k\ell}^{(n)}) \tag{3.14}$$

where $t_{ij}^{(n)}$ and $\Delta_{ij}^{(n)}$ are defined as in (3.5). Observe that the summation condition ($i<k$, $j>\ell$) implies that only the values of $\psi_{s,s'}$ on the set $s\curlywedge s'$ affect the integral.

By taking $\mu(ds)$ and $\nu(ds)$ alternatively as ds and $W(ds)$ we get four different types of integrals for (3.12). The case of

$$\int \psi_{s,s'}\,ds\,ds'$$

is uninteresting because it is representable as an ordinary Lebesgue integral. The case of

$$\int \psi_{s,s'}W(ds)W(ds')$$

gives us the type-II stochastic integral. The remaining two possibilities

$$\int \psi_{s,s'}W(ds)ds', \quad \int \psi_{s,s'}\,ds\,W(ds')$$

will be called mixed integrals and they play an essential role in the stochastic calculus of two-parameter martingales.

4. DIFFERENTIATION FORMULAS AND WEAK MARTINGALES

The next natural question that arises concerns transformation rule for stochastic integrals, i.e., generalization of the Ito differentiation formula. This turned out to be rather difficult, but a restricted version of the differentiation formula was obtained in [1]. Suppose that M_t is an $\mathscr{F}_{Wt}$-martingale which is also a type-I stochastic integral, i.e., of the form (3.9). Let $F(u,t)$, $u \in R$, $t \in R_+^2$, be a suitably differentiable function such that

$$X_t = F(M_t,t) \tag{4.1}$$

is again an $\mathscr{F}_{Wt}$-martingale. Then the representation of X in terms of stochastic integrals is given by

$$X_t = X_0 + \int_{A_t} F'(M_s,s)M(ds) + \int_{A_t\times A_t} F''(M_{s\vee s'},s\vee s')M(ds)M(ds') \tag{4.2}$$

where F' and F'' denote differentiations of F with respect to the first variable, $s\vee s' = (\max(s_1,s_1'), \max(s_2,s_2'))$, and $M(ds) = \phi_s W(ds)$.

As a transformation formula (4.2) is severely limited. It deals only with transformations of type-I stochastic integrals which result in martingales. The general question is the following:

Suppose that X_t is of the form

$$(4.3)\qquad X_t = \int_{A_t} \theta_s ds + \int_{A_t} \phi_s W(ds) + \int_{A_t \times A_t} \psi_{s,s'} W(ds)W(ds')$$

and $F(x,t)$ is a suitable differentiable function. Can $F(X_t,t)$ again be expressed as the sum of three such integrals?

The answer is once again "no". This means that the two types of stochastic integrals together with the Lebesgue integral are still not enough to give us a complete stochastic calculus.

To get a complete stochastic calculus we need the mixed integrals defined at the end of the last section. The substance of the differentiation formula is that if X_t is of the form

$$(4.4)\qquad X_t = \int_{A_t} \theta_s ds + \int_{A_t} \phi_s W(ds) + \int_{A_t \times A_t} \psi_{s,s'} W(ds)W(ds')$$

$$+ \int_{A_t \times A_t} f_{s,s'} W(ds)ds' + \int_{A_t \times A_t} g_{s,s'} ds W(ds')$$

then $F(X_t,t)$ is again the sum of five such integrals. The full differentiation formula is rather complicated and will not be given here. (See [2])

Stochastic integrals of the two types are martingales. The mixed integrals clearly cannot be martingales, because there would be no need to introduce them otherwise. However, they are <u>weak martingales</u>. [3] A process $\{X_t, \mathscr{F}_t, t \in T\}$ is said to be a weak martingale if X is $\mathscr{F}$-adapted and

$$(4.5)\qquad E[X(\Delta_t) | \mathscr{F}_t] = 0$$

where

$$(4.6)\qquad X(\Delta_t) = X_{(t_1+\Delta_1, t_2+\Delta_2)} - X_{(t_1+\Delta_1, t_2)} - X_{(t_1, t_2+\Delta_2)} - X_t$$

All martingales are also weak martingales, but not conversely. We note that (4.5) is a natural generalization of the alternative definition for martingales in one dimension given by

$$E[X(t+\Delta) - X(t) | \mathscr{F}_t] = 0$$

The substance of the general differentiation formula can now be taken to mean that a suitably differentiable function of a weak martingale plus a Lebesgue integrable is again a weak martingale plus a Lebesgue integral.

5. LIKELIHOOD RATIO FORMULAS

Let us return to the observation equation (1.1) with which we began our discussion. Integrating both sides of (1.1) we get a new observation equation

$$(5.1) \qquad Y_t = \int_{A_t} X_s ds + W_t, \qquad t \in T$$

where we now consider Y as the observation and W the noise process. Let $\mathscr{F}_t$ denote the σ-field generated by $\{X_s, W_s, s \in A_t\}$ and let $\mathscr{F}_{yt}$ denote the sub-σ-field of $\mathscr{F}_t$ generated by $\{Y_s, s \in A_t\}$. If $\mathscr{P}$ denotes the probability measure, then our earlier assumption that the noise is white and Gaussian is equivalent to say that W_t is a Wiener process under $\mathscr{P}$. Let us also assume that the noise is independent of the signal then W is a Wiener process with respect to $(\mathscr{P}, \{\mathscr{F}_t\})$, and it was shown in [4] that there exists a new probability measure $\mathscr{P}_0$ such that Y_t is a Wiener process with respect to $(\mathscr{P}_0, \{\mathscr{F}_t\})$. The density (Radon-Nikodym derivative) of $\mathscr{P}$ with respect to $\mathscr{P}_0$ is given by

$$(5.2) \qquad \frac{d\mathscr{P}}{d\mathscr{P}_0} = \exp\{\int_T X_s Y(ds) - \frac{1}{2}\int_T X_s^2 ds\}$$

Observe that since Y is a Wiener process, under $\mathscr{P}_0$ the first integral is a type-I stochastic integral.

The likelihood ratio L_t defined by

$$(5.3) \qquad L_t = E_0\{\frac{d\mathscr{P}}{d\mathscr{P}_0} \Big| \mathscr{F}_{yt}\}$$

plays a prominent role in both hypothesis testing (detection) and estimation problems, and it is important to find explicit expressions for it. First, it can be observed immediately that by virtue of its definition L_t must be a $(\mathscr{P}_0, \{\mathscr{F}_{yt}\})$ martingale. Since Y_t is a $(\mathscr{P}_0, \{\mathscr{F}_{yt}\})$ Wiener process, this means that we can write L_t in the form

$$(5.4) \qquad L_t = 1 + \int_{A_t} \phi_s Y(ds) + \int_{A_t \times A_t} \psi_{s,s'} Y(ds) Y(ds')$$

The integrands ϕ and ψ were identified in [4], and (5.4) can be rewritten as

$$(5.5) \qquad L_t = 1 + \int_{A_t} \hat{X}(s|s) L_s Y(ds)$$

$$+ \int_{A_t \times A_t} [\rho(s,s'|s \vee s') + \hat{X}(s|s \vee s')\hat{X}(s'|s \vee s')] L_{s \vee s'} Y(ds) Y(ds')$$

where

(5.6) $\hat{X}(s|s) = E[X_s | \mathscr{F}_{ys}]$

and

(5.7) $\rho(s,s'|s\vee s') = \mathrm{Cov}(X_s X_{s'} | \mathscr{F}_{ys\vee s'})$

Equation (5.4) shows that L_t is expressible in terms of (5.6) and (5.7) which are conditional moments of X given the observation. Unlike the one-dimensional case, the likelihood ratio now depends on the second as well as the first moment. The next obvious step was to treat (5.5) as an integral equation in L and solve it to obtain an explicit expression for L_t in terms of $\hat{X}$ and ρ. This proved to be rather difficult and had to await further developments in the stochastic calculus. The desired expression is derived in [5] and has the form

$$(5.8)\quad L_t = \exp\Big\{\int_{A_t} \hat{X}(s|s)Y(ds) - \frac{1}{2}\int_{A_t} \hat{X}^2(s|s)ds + \int_{A_t\times A_t} \rho(s,s'|s\vee s')[Y(ds)-\hat{X}(s|s\vee s')ds][Y(ds')-\hat{X}(s'|s\vee s')ds'] - \frac{1}{2}\int_{A_t\times A_t} \rho^2(s,s'|s\vee s')ds\,ds'\Big\}$$

Equation (5.8) is an exceedingly interesting formula, and its form could not have been predicted from its one-dimensional counterpart.

6. INNOVATIONS AND RECURSIVE FILTERING

Consider the observation equation (5.1) once again, and pose the following question:

Suppose that Z_t is a martingale with respect to $(\mathscr{P},\{\mathscr{F}_{yt}\})$. What is the general form for Z_t?

This turns out to be a difficult question and a satisfactory answer is not yet known. Although this question may appear to be the most natural generalization of the innovations representation problem in one dimension, it is actually not. In terms of the form of the answer and the usefulness of the answer, the most natural generalization of the innovation problem is the following.

What is the general form of a <u>weak</u> martingale with respect to $(\mathscr{P},\{\mathscr{F}_{yt}\})$?

The answer is that if Z_t is a $(\mathscr{P},\{\mathscr{F}_{yt}\})$ weak martingale then it must be of the form

$$(6.1)\quad Z_t = Z_0 + \int_{A_t} \phi_s \hat{Y}(ds|s) + \int_{A_t\times A_t} f_{s,s'}\hat{Y}(ds|s\vee s')ds' + \int_{A_t\times A_t} g_{s,s'}ds\hat{Y}(ds'|s\vee s')$$

$$+\int_{A_t\times A_t}\psi_{s,s'}[\hat{Y}(ds|s\vee s')\hat{Y}(ds'|s\vee s')-\rho(s,s'|s\vee s')dsds']$$

where $\hat{Y}$ is defined by

$$(6.2)\qquad \hat{Y}(ds|t) = Y(ds)-\hat{X}(s|t)ds$$

Equation (6.1) leads rather quickly to a recursive formula for computing $\hat{X}$, when the signal X satisfies a modelling equation of an appropriate type. Without attempting to get the most general result possible, assume that X is a Gaussian process which satisfies the differential equation

$$(6.3)\quad \frac{\partial^2}{\partial t_1\partial t_2}X_t = \alpha(t)\frac{\partial}{\partial t_1}X_t+\beta(t)\frac{\partial}{\partial t_2}X_t+\gamma(t)X_t+\zeta_t$$

where ξ_t is again a white Gaussian noise. We can rewrite it as

$$(6.4)\qquad d_{t_1}d_{t_2}X_t-\alpha(t)d_{t_1}X_t dt_2-\beta(t)dt_1 d_{t_2}X_t-\gamma(t)X_t dt_1 dt_2 = V(dt)$$

where V is a Wiener process. Now, it is easily shown that if we define a process M by

$$(6.5)\qquad d_{t_1}d_{t_2}\hat{X}(t|t)-\alpha(t)d_{t_1}\hat{X}(t|t)dt_2-\beta(t)dt_1 d_{t_2}\hat{X}(t|t)$$
$$=\gamma(t)\hat{X}(t|t)dt_1dt_2 = M(dt)$$

then M must be a weak martingale with respect to $(\mathscr{P}, \mathscr{F}_{yt})$. This means that M_t must be of the form given by (6.1). Furthermore, X is Gaussian and the observation equation is linear. It follows that $\hat{X}(t|t)$ must be linear in Y, and for M the last term in (6.1) must vanish. It also means that the integrands in the first three integrals of (6.1) must be deterministic functions for M. Indeed, these integrands can all be shown to be expressible in terms of the covariance function ρ, and M(dt) can be expressed as

$$(6.6)\qquad M(dt) = \frac{1}{N_0}\rho(t,t|t)\hat{Y}(dt|t)$$
$$+\frac{1}{N_0}\int_0^{t_2} d_{t_2}\rho(t_1,t_2;t_1,s_2|t)\hat{Y}(dt_1ds_2|t)$$
$$+\frac{1}{N_0}\int_0^{t_1} d_{t_1}\rho(t_1,t_2;s_1,t_2|t)Y(ds_1dt_2|t)$$

Equations (6.5) and (6.6) together with the definition (6.2) for Y reveal the nature of recursion for $\hat{X}(t|t)$. They show that $d_{t_1}d_{t_2}X(t|t)$ depends not only on $\hat{X}(t|t)$ but also on $\hat{X}(s|t)$ for s on the boundary ∂A_t of the rectangle A_t. Instead of (6.5), a more interesting recursion is in terms of the boundary data

$$\hat{X}(\partial t) = \{\hat{X}(s|t), \quad s \in \partial A_t\}$$

If $t' \succ t$, then $\hat{X}(\partial t')$ can be computed from $\hat{X}(\partial t)$ and the observation in the area between A_t and $A_{t'}$. The details can be found in [6].

REFERENCES

1. E. Wong and M. Zakai, "Martingales and stochastic integrals for processes with a multidimensional parameter," Zeit. Wahrscheinlichkeitstheorie, Vol. 29, pp. 109-122, 1974.
2. E. Wong and M. Zakai, "Differentiation formulas for stochastic integrals in the plane," Stochastic Processes and Their Applications, In press.
3. R. Cairoli and J.B. Walsh, "Stochastic integrals in the plane," Acta Math., Vol. 134, pp. 111-183, 1975.
4. E. Wong, "A likelihood ratio formula for two-dimensional random fields," IEEE Trans. Information Theory, Vol. IT-20, pp. 418-422.
5. E. Wong and M. Zakai, "Likelihood ratios and transformation of probability associated with two-parameter Wiener processes," Zeit. Wahrscheinlichkeitstheorie. In press.
6. E. Wong, "Recursive causal linear filtering for two-dimensional random fields," IEEE Trans. Information Theory. In press.