CHAPTER 11

Probability Density Functionals and Reproducing Kernel Hilbert Spaces*

Emanuel Parzen, Stanford University

ABSTRACT

The extraction, detection, and prediction of signals in the presence of noise are among the central problems of statistical communication theory. Over the past few years I have sought to develop an approach to those problems that would simultaneously apply to stationary or nonstationary, discrete parameter or continuous parameter, and univariate or multivariate time series and would distinguish between their statistical and analytical aspects. In particular, they would clarify the role played by various widely employed analytical techniques (such as the Wiener-Hopf equation and eigenfunction expansions).

In the development of this approach, two basic concepts are used: the notion of the probability density functional of a time series and the notion of a reproducing kernel Hilbert space. The purpose of this chapter is to sketch the relation between these concepts.

1. THE PROBABILITY DENSITY FUNCTIONAL OF A NORMAL TIME SERIES

Let $[S(t), t \in T]$ and $[N(t), t \in T]$ be time series, called, respectively, the signal process and the noise process. Let Ω be the space of all real-valued functions on T. Let P_N and P_{S+N} be probability measures defined on the measurable subsets B of Ω by

$$P_N[B] = \text{prob}\ \{[N(t),\ t \in T] \in B\} \tag{1}$$

$$P_{S+N}[B] = \text{prob}\ \{[S(t) + N(t),\ t \in T] \in B\}. \tag{2}$$

We are trying to determine, if it exists, a function p on Ω with the property that

$$P_{S+N}[B] = \int_B p\, dP_N. \tag{3}$$

* Prepared under contract Nonr 3440(00) for the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The function p may be called the *probability density functional* of P_{S+N} with respect to P_N in order to emphasize that its argument is a function $[X(t), t \in T]$. It is also denoted $p[X(t), t \in T]$ and called the probability density functional of the signal-plus-noise process

$$X(t) = S(t) + N(t), t \in T, \tag{4}$$

with respect to the noise process $[N(t), t \in T]$. The function p is often written symbolically as a derivative,

$$p = \frac{dP_{S+N}}{dP_N} \tag{5}$$

and called the Radon-Nikodym derivative of P_{S+N} with respect to P_N [see Halmos (1950), p. 329].

A necessary and sufficient condition that the probability density (5) exist is that P_{S+N} be *absolutely continuous* with respect to P_N in the sense that, for every measurable subset A of Ω,

$$P_N[A] = 0 \text{ implies } P_{S+N}[A] = 0. \tag{6}$$

In order that P_{S+N} be *not* absolutely continuous with respect to P_N, it is necessary and sufficient that there exist a set A such that

$$P_N[A] = 0 \text{ and } P_{S+N}[A] > 0. \tag{7}$$

The probability measures P_N and P_{S+N} are said to be *orthogonal* if there exists a set A such that

$$P_N[A] = 0 \text{ and } P_{S+N}[A] = 1. \tag{8}$$

We can regard (8) as an extreme case of being not absolutely continuous.

The notion of orthogonality derives its importance from detection theory (the theory of testing hypotheses). The simple hypotheses

$$H_0: X(\cdot) = N(\cdot)$$
$$H_1: X(\cdot) = S(\cdot) + N(\cdot)$$

are said to be *perfectly detectable* if there exists a set A such that

$$\begin{aligned} P_N[A] &= \text{prob}\,\{[X(t), t \in T] \in A | H_0\} = 0 \\ P_{S+N}[A] &= \text{prob}\,\{[X(t), t \in T] \in A | H_1\} = 1. \end{aligned} \tag{9}$$

Clearly, the hypotheses H_0 and H_1 are perfectly detectable if and only if P_N and P_{S+N} are orthogonal.

Given the probability measures P_N and P_{S+N}, the following questions arise:

1. Determine whether P_N and P_{S+N} are orthogonal.
2. Determine whether P_{S+N} is absolutely continuous with respect to P_N.
3. Determine the Radon-Nikodym derivative (5) if it exists.

To answer these questions, the natural way to proceed is to approximate the infinite dimensional case by finite dimensional cases. For any finite subset

$$T' = (t_1, \cdots, t_n) \text{ of } T \tag{10}$$

let $P_{N,T'}$ and $P_{S+N,T'}$ denote the probability distributions of $[X(t),\ t \in T']$ under P_N and P_{S+N}, respectively. *Assume that $P_{S+N,T'}$ is absolutely continuous with respect to $P_{N,T'}$*, with Radon-Nikodym derivative denoted

$$p_{T'} = \frac{dP_{S+N,T'}}{dP_{N,T'}}. \tag{11}$$

The *divergence* between P_{S+N} and P_N on the basis of having observed $[X(t),\ t \in T']$ is defined by

$$J_{T'} = E_{S+N}(\log p_{T'}) - E_N(\log p_{T'}). \tag{12}$$

Using the theory of martingales, it may be shown that

$$0 \leqslant J_{T'} \leqslant J_{T''} \quad \text{if} \quad T' \subset T''. \tag{13}$$

Consequently, the limit

$$J_T = \lim_{T' \to T} J_{T'} \tag{14}$$

exists and is finite or infinite. Further, it may be shown [see Hajek (1958)] that (*a*) if $J_T < \infty$, then P_{S+N} is absolutely continuous with respect to P_N and

$$p = \frac{dP_{S+N}}{dP_N} = \lim_{T' \to T} p_{T'}; \tag{15}$$

(*b*) if $J_T = \infty$, and both the time series $[N(t),\ t \in T]$ and $[S(t) + N(t),\ t \in T]$ are normal, then P_{S+N} and P_N are orthogonal.

We next apply these criteria under the following assumptions.

The noise process $[N(t),\ t \in T]$ is a normal process with zero means and covariance kernel

$$K(s, t) = E[N(s)\, N(t)], \tag{16}$$

which is positive definite in the sense that for every finite subset $T' = \{t_1, \cdots, t_n\}$ of T the covariance matrix

$$K_{T'} = [K(t_i, t_j)] = \begin{bmatrix} K(t_1, t_1) & \cdots & K(t_1, t_n) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ K(t_n, t_1) & \cdots & K(t_n, t_n) \end{bmatrix} \tag{17}$$

is nonsingular, with inverse matrix denoted

$$K_{T'}^{-1} = [K^{-1}(t_i, t_j)]. \tag{18}$$

(It should be noted that the assumption of positive definiteness is made only for mathematical convenience in the present exposition; it can be omitted.)

In regard to the signal process, two cases are of most interest:

1. *Sure signal case.* $[S(t),\, t \in T]$ is a nonrandom function.
2. *Stochastic signal case.* $[S(t),\, t \in T]$ is a normal time series, independent of the noise process, with zero means and positive definite covariance kernel

$$R(s, t) = E[S(s)\, S(t)]. \tag{19}$$

To employ the criterion (15), we first need to compute the divergence $J_{T'}$, defined by (12). In this section we consider the sure signal case; the stochastic signal case is considered in Section 3.

In the sure signal case

$$\log p_{T'} = (X, S)_{K,T'} - \tfrac{1}{2}(S, S)_{K,T'} \tag{20}$$

where we define for any functions f and g on T

$$(f, g)_{K,T'} = \sum_{s,t \in T'} f(s) K^{-1}(s, t)\, g(t). \tag{21}$$

Consequently

$$J_{T'} = E_{S+N}[(X, S)_{K,T'}] - E_N[(X, S)_{K,T'}] = (S, S)_{K,T'} \tag{22}$$

and

$$J_T < \infty \text{ if and only if } \lim_{T' \to T} (S, S)_{K,T'} < \infty. \tag{23}$$

In words, in the sure signal case, P_{S+N} is absolutely continuous with respect to P_N if and only if $(S, S)_{K,T'}$ approaches a limit as T' tends to T. Fortunately it is possible to characterize those functions $S(\cdot)$ that have this property. To do so, we introduce the notion of a reproducing kernel Hilbert space.

2. REPRODUCING KERNEL HILBERT SPACES

Let $K(s, t)$ be the covariance kernel of a time series $[X(t),\, t \in T]$. For each t in T, let $K(\cdot, t)$ be the function on T whose value at s in T is equal to $K(s, t)$. It may be shown [see Aronszajn (1950)] that there exists a unique Hilbert space, denoted $H(K; T)$, with the following properties:

1. The members of $H(K; T)$ are real-valued functions on T [if $K(s, t)$ were complex-valued, they would be complex-valued functions].
2. For every t in T

$$K(\cdot, t) \in H(K; T). \tag{I}$$

3. For every t in T and f in $H(K; T)$

$$f(t) = (f, K(\cdot, t))_{K,T}, \tag{II}$$

where the inner product between two functions f and g in $H(K; T)$ is written $(f, g)_{K,T}$.

Example 1. Suppose $T = (1, 2, \ldots, n)$ for some positive integer n and that the covariance kernel K is given by a symmetric positive definite matrix $[K_{ij}]$ with inverse $[K^{ij}]$. The corresponding reproducing kernel space $H(K; T)$

consists of all n-dimensional vectors $\mathbf{f} = (f_1, \cdots, f_n)$ with inner product

$$(\mathbf{f}, \mathbf{g})_{K,T} = \sum_{s,t=1}^{n} f_s K^{st} g_t. \tag{24}$$

To prove (24) we need only to verify that the reproducing property holds for $u = 1, \cdots, n$

$$(\mathbf{f}, K_{\cdot u})_{K,T} = \sum_{s,t=1}^{n} f_s K^{st} K_{tu} = \sum_{s=1}^{n} f_s\, \delta(s, u) = f_u.$$

The inner product may also be written as a ratio of determinants:

$$(\mathbf{f}, \mathbf{g})_{K,T} = -\begin{vmatrix} K_{11} & \cdots & K_{1n} & f_1 \\ \cdot & & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ K_{n1} & \cdots & K_{nn} & f_n \\ g_1 & \cdots & g_n & 0 \end{vmatrix} \div \begin{vmatrix} K_{11} & \cdots & K_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ K_{n1} & \cdots & K_{nn} \end{vmatrix}. \tag{25}$$

To prove (25), we again need only to verify the reproducing property. When the covariance matrix $\mathbf{K}$ is singular, we may define the corresponding reproducing kernel inner product in terms of the *pseudo-inverse* of the matrix $\mathbf{K}$.

Example 2. Let $T = [t: a \leqslant t \leqslant b]$ and let $[N(t), a \leqslant t \leqslant b]$ be the Wiener process; that is, it has independent increments and covariance function

$$K(s, t) = \sigma^2 \min(s, t) \tag{26}$$

for some parameter σ^2. Consider the Hilbert spaces $H(K; T)$ consisting of all functions f on $a \leqslant t \leqslant b$ of the form

$$f(t) = f(a) + \int^{t} f'(u)\, du \tag{27}$$

for some square integrable measurable function f' on $a \leqslant t \leqslant b$ [which can be called the L_2-derivative of f], with inner product defined by

$$(f, g)_{K,T} = \frac{1}{\sigma^2}\left[\frac{1}{a} f(a)\, g(a) + \int_a^b f'(u)\, g'(u)\, du\right]. \tag{28}$$

If we define

$$\begin{aligned} I_t^{(u)} &= 1 \quad \text{if} \quad a \leqslant u \leqslant t \\ &= 0 \quad \text{if} \quad t < u \leqslant b, \end{aligned} \tag{29}$$

we may rewrite (27):

$$f(t) = f(a) + \int_a^b f'(u)\, I_t(u)\, du.$$

Now the covariance kernel $K(s, t)$ may be represented as

$$K(s, t) = \sigma^2 a + \sigma^2 \int_a^b I_s(u)\, I_t(u)\, du.$$

Therefore, for each t in T, $K(\cdot, t)$ belongs to $H(K; T)$ with L^2 derivative

$$\frac{d}{ds} K(s, t) = \sigma^2 I_t(s).$$

Further,

$$(f, K(\cdot, t))_{K,T} = \frac{1}{\sigma^2}\left[\frac{1}{a} f(a)\sigma^2 a + \int_a^b f'(u)\sigma^2\, I_t(u)\, du\right]$$

$$= f(a) + \int_a^t f'(u)\, du = f(t).$$

Thus we see that the reproducing kernel Hilbert space corresponding to the covariance kernel (26) consists of all L_2-differentiable functions on T with inner product given by (28).

The relevance of the theory of reproducing kernel Hilbert spaces to the theory of probability density functionals derives from the following fact: it may be shown (using martingale theory) that

$$\lim_{T' \to T} (S, S)_{K,T'} < \infty \quad \text{if and only if} \quad S \in H(K; T). \tag{30}$$

Further, if $S \in H(K; T)$, then

$$\lim_{T' \to T} (S, S)_{K,T'} = (S, S)_{K,T}. \tag{31}$$

It follows in the sure signal case that P_{S+N} is absolutely continuous with respect to P_N if and only if the signal function $[S(t), t \in T]$ belongs to the reproducing kernel Hilbert space $H(K; T)$ corresponding to the covariance kernel K of the noise process $[X(t), t \in T]$. If $S \in H(K; T)$, then the probability density functional is given by

$$p[X(t), t \in T] = \frac{dP_{S+N}}{dP_N} = \exp\left[(X, S)_{K,T} - \tfrac{1}{2}(S, S)_{K,T}\right] \tag{32}$$

where by $(X, S)_{K,T}$ we mean the limit (in the sense both of convergence with probability one and convergence in quadratic mean)

$$(X, S)_{K,T} = \lim_{T' \to T} (X, S)_{K,T'}. \tag{33}$$

It should be emphasized that although we use inner product notation to write $(X, S)_{K,T}$ this is not a true inner product between two elements in a Hilbert space, since the sample function $[X(t), t \in T]$ does not belong to $H(K)$; that is,

$$\lim_{T' \to T} (X, X)_{K,T'} \text{ is infinite with probability one.} \tag{34}$$

In practice, it will be clear how to define $(X, S)_{K,T}$ by suitably modifying the expression for the inner product between two functions in $H(K)$. Thus for the covariance kernel given by (26), instead of the expression

$$(X, S)_{K,T} = \frac{1}{\sigma^2}\left[\frac{1}{a} X(a)\, S(a) + \int_a^b S'(u)\, X'(u)\, du\right]$$

suggested by (28), we may show that

$$(X, S)_{K,T} = \frac{1}{\sigma^2}\left[\frac{1}{a} X(a)\, S(a) + \int_a^b S'(u)\, dX(u)\right].$$

There is a variety of ways in which one can determine whether a function S belongs to a reproducing kernel Hilbert space $H(K; T)$ and compute the norm $(S, S)_{K,T}$ and the random variable $(X, S)_{K,T}$. These are discussed elsewhere [see Parzen (1961)].

However, certain general principles deserve to be stated at this point.

Roughly speaking, a function $g(\cdot)$ belongs to a reproducing kernel Hilbert space $H(K; T)$ only if it is at least as "smooth" as the functions $K(\cdot, t)$, since every function g in $H(K; T)$ is either a linear combination

$$g(\cdot) = \sum_{i=1}^{n} c_i K(\cdot, t_i)$$

or a limit of such linear combinations. For example, if T is an interval and K is continuous on $T \otimes T$, then every function in $H(K; T)$ is continuous; if K is twice differentiable on $T \otimes T$, then every function in $H(K, T)$ is differentiable.

We are led to the following heuristic conclusion: *In order that a signal not be perfectly detectable in the presence of a noise, it is necessary and sufficient that the signal be as smooth as the noise.* In the case of a sure signal the signal is as smooth as the noise if and only if $S \in H(K; T)$, where K is the covariance kernel of the noise. In the case of stochastic signals the signal is as smooth as the noise if $S \in H(K; T)$ for almost all sample functions of the signal process: a rigorous formulation of this assertion is given in Section 3.

A basic tool in the analytical evaluation of a reproducing kernel inner product is provided by the following theorem.

Integral representation theorem

Let K be a covariance kernel. If (a) a measurable space $(Q, \mathbf{B}, \mu)$ exists and (b) in the Hilbert space of all $\mathbf{B}$-measurable real-valued functions on Q satisfying

$$(f, f)_\mu = \int_Q f^2\, d\mu < \infty \tag{35}$$

there exists a family $[f(t), t \in T]$ of functions satisfying

$$K(s, t) = (f(s), f(t))_\mu = \int_Q f(s)\, f(t)\, d\mu, \tag{36}$$

then the reproducing kernel Hilbert space $H(K; T)$ consists of all functions

g on T, which may be represented as

$$g(t) = \int_Q g^* f(t)\, d\mu, \tag{37}$$

for some unique function g^* in the Hilbert subspace $L[f(t), t \in T]$ of $L_2(Q, \mathbf{B}, \mu)$ spanned by the family of functions $[f(t), t \in T]$. The norm of g is given by

$$(g, g)_{K,T} = (g^*, g^*)_\mu \tag{38}$$

If $[f(t), t \in T]$ spans $L_2(Q, \mathbf{B}, \mu)$, then $X(t)$ may be represented as a stochastic integral with respect to an orthogonal random set function $[Z(B), B \in \mathbf{B}]$ with covariance kernel μ:

$$X(t) = \int_Q f(t)\, dZ \tag{39}$$

$$E[Z(B_1)\, Z(B_2)] = \mu(B_1 B_2). \tag{40}$$

Further

$$(X, g)_{K,T} = \int_Q g^*\, dZ. \tag{41}$$

As an immediate consequence of the integral representation theorem one obtains the following example.

Example 3. *Stationary noise process.* Let $T = [t\colon -\infty < t < \infty]$ and let $[X(t), -\infty < t < \infty]$ be a stationary time series with spectral density function $f(\omega)$ so that

$$K(s, t) = \int_{-\infty}^{\infty} e^{i\omega(s-t)} f(\omega)\, d\omega. \tag{42}$$

Then $H(K; T)$ consists of all functions g on T of the form

$$g(t) = \int_{-\infty}^{\infty} g^*(\omega) e^{i\omega t} f(\omega)\, d\omega \tag{43}$$

for which the norm

$$\|g\|^2_{K,T} = \int_{-\infty}^{\infty} |g^*(\omega)|^2 f(\omega)\, d\omega \tag{44}$$

is finite. The corresponding random variable $(X, g)_{K,T}$ can be expressed in terms of the spectral representation of $X(\cdot)$. If

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega}\, dZ(\omega), \tag{45}$$

then

$$(X, g)_{K,T} = \int_{-\infty}^{\infty} g^*(\omega)\, dZ(\omega). \tag{46}$$

Assume that the spectral density function $f(\omega)$ is uniformly bounded. Then $g(t)$, being the Fourier transform of the square integrable function $g^*(\omega) f(\omega)$, is square integrable. Let

$$G(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\omega} g(t)\, dt. \tag{47}$$

Then

$$g^*(\omega) f(\omega) = G(\omega). \tag{48}$$

Assume now that $f(\omega)$ never vanishes. We may then write

$$(g, g)_{K,T} = \int_{-\infty}^{\infty} \left| \frac{G(\omega)}{f(\omega)} \right|^2 f(\omega)\, d\omega = \int_{-\infty}^{\infty} |G(\omega)|^2 \frac{1}{f(\omega)}\, d\omega \tag{49}$$

$$(X, g)_{K,T} = \int_{-\infty}^{\infty} \frac{G(\omega)}{f(\omega)}\, dZ(\omega). \tag{50}$$

To sum up, in the case of a stationary process whose spectral density function is uniformly bounded and never vanishes, the reproducing kernel Hilbert space $H(K; T)$ for $T = (-\infty < t < \infty)$ consists of all space integrable functions $g(t)$ whose Fourier transforms $G(\omega)$ are such that

$$\int_{-\infty}^{\infty} |G(\omega)|^2 \frac{1}{f(\omega)}\, d\omega < \infty. \tag{51}$$

If $f(\omega)$ vanishes, a similar conclusion holds. Let $N = [\omega: f(\omega) = 0]$ and $N^c = [\omega: f(\omega) > 0]$. Then $H(K; T)$ consists of all square integrable functions $g(t)$ whose Fourier transforms $G(\omega)$ vanish on N and such that

$$\int_{N^c} |G(\omega)|^2 \frac{1}{f(\omega)}\, d\omega < \infty.$$

The foregoing results are easily extended to multiple time series $[X_\alpha(t), -\infty < t < \infty, \alpha = 1, 2, \cdots, M]$. Assume that for $\alpha, \beta = 1, 2, \cdots, M$ and $-\infty < s, t < \infty$,

$$K_{\alpha,\beta}(s, t) = E[X_\alpha(s)\, X_\beta(t)] = \int_{-\infty}^{\infty} e^{i\omega(s-t)} f_{\alpha,\beta}(\omega)\, d\omega. \tag{52}$$

Then $H(K; T)$ consists of all functions $g_\alpha(t)$ on

$$T = [(\alpha, t): \alpha = 1, \cdots, M, -\infty < t < \infty], \tag{53}$$

satisfying the condition

$$\sum_{\alpha=1}^{M} \int_{-\infty}^{\infty} g_\alpha^2(t)\, dt < \infty, \tag{54}$$

such that

$$\int_{-\infty}^{\infty} \left[\sum_{\alpha,\beta=1}^{M} G_\alpha(\omega)\, f^{\alpha\beta}(\omega)\, \overline{G_\beta(\omega)} \right] d\omega < \infty, \tag{55}$$

where $\bar{z}$ denotes the complex conjugate of the complex number z,

$$G_\alpha(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t}\, g_\alpha(t)\, dt, \tag{56}$$

and $[f^{\alpha\beta}(\omega)]$ is the inverse of the matrix $[f_{\alpha\beta}(\omega)]$. Then $(g, g)_{K,T}$ is given by the expression in (55) and

$$(X, g)_{K,T} = \sum_{\alpha,\beta=1}^{M} \int_{-\infty}^{\infty} \overline{G_\alpha(\omega)}\, f^{\alpha\beta}(\omega)\, dZ_\beta(\omega), \tag{57}$$

where

$$X_\alpha(t) = \int_{-\infty}^{\infty} e^{it\omega}\, dZ_\alpha(\omega). \tag{58}$$

Direct Product Hilbert Spaces

The notion of a direct product space plays an important part in our considerations. Given two function spaces G_1 and G_2, consisting of functions defined on T_1 and T_2, respectively, their direct product space, denoted $G_1 \otimes G_2$, is the Hilbert space completion of the set of functions g on $T_1 \otimes T_2$ of the form

$$g(t_1, t_2) = g_1(t_1)\, g_2(t_2), \tag{59}$$

where $g_1 \in G_1$ and $g_2 \in G_2$. The norm of a function in $G_1 \otimes G_2$ of the form (59) is defined by

$$\|g\|^2_{G_1\otimes G_2} = \|g_1\|^2_{G_1}\|g_2\|^2_{G_2}. \tag{60}$$

The function g defined by (59) is on occasion denoted by $g_1 \otimes g_2$.

It should be noted that if G_1 and G_2 are reproducing kernel Hilbert spaces, with respective reproducing kernels K_1 and K_2 defined on $T \otimes T$, then $G_1 \otimes G_2$ is a reproducing kernel Hilbert space with kernel $K_1 \otimes K_2$, where $K_1 \otimes K_2$ is a function of four real variables defined by

$$K_1 \otimes K_2(s_1, s_2, t_1, t_2) = K_1(s_1, t_1)\, K_2(s_2, t_2) \tag{61}$$

and

$$(g, K_1 \otimes K_2(\cdot, \cdot, t_1, t_2))_{G_1\otimes G_2} = g(t_1, t_2). \tag{62}$$

When $G_1 = G_2 = L_2(T, \mathbf{B}, \mu)$, $G_1 \otimes G_2$ consists of all ($\mathbf{B} \otimes \mathbf{B}$-measurable) functions g on $T \otimes T$ such that

$$\|g\|^2_{G_1\otimes G_2} = \int_T \int_T g^2(s, t)\, \mu(ds)\, \mu(dt) < \infty \tag{63}$$

If G_1 and G_2 are equal to the reproducing kernel Hilbert space consisting of all L_2-differentiable functions on the interval $(t\colon a \leqslant t \leqslant b)$ with norm squared

$$\|g\|^2_{G_1} = \frac{1}{a} g_1^2(a) + \int_a^b |g_1'(t)|^2\, dt, \tag{64}$$

then $G_1 \otimes G_2$ is a reproducing kernel Hilbert space with norm squared

$$\begin{aligned} \|g\|^2_{G_1\otimes G_2} = \frac{1}{a^2} g^2(a, a) &+ \frac{1}{a}\int_a^b \left|\frac{\partial}{\partial s} g(s, a)\right|^2 ds \\ &+ \frac{1}{a}\int_a^b \left|\frac{\partial}{\partial t} g(a, t)\right|^2 dt \\ &+ \int_a^b \int_a^b \left|\frac{\partial}{\partial s}\frac{\partial}{\partial t} g(s, t)\right|^2 ds\, dt. \end{aligned} \tag{65}$$

3. STOCHASTIC SIGNAL CASE

In this section we shall determine conditions for the existence of the probability density functional (6) in the stochastic signal case described before (19).

We shall prove below that P_{S+N} is absolutely continuous with respect to P_N if and only if

$$\|R\|_{H(K)\otimes H(K+R)} < \infty. \tag{66}$$

It may be shown that a sufficient condition for (66) to hold is that

$$\|R\|_{H(K)\otimes H(K)} < \infty. \tag{67}$$

In practice, the condition we shall attempt to verify is (67). Consequently, before proving that (66) is necessary and sufficient for $p = dP_{S+N}/dP_N$ to exist, let us show directly that (67) is a sufficient condition for p to exist and let us obtain an explicit formula for p.

It may be shown that if (67) holds, then the signal process $[S(t), t \in T]$ may be written

$$S(t) = \sum_{\nu=1}^{\infty} \eta_\nu \Phi_\nu(t), \tag{68}$$

where (*a*) $[\eta_\nu]$ is a sequence of random variables satisfying

$$E(\eta_\alpha, \eta_\beta) = \delta(\alpha, \beta)\lambda_\alpha \tag{69}$$

for a suitable sequence $[\lambda_\nu]$, and (*b*) $[\Phi_\nu]$ is a sequence of functions in $H(K)$ satisfying

$$(\Phi_\alpha, \Phi_\beta)_{H(K)} = \delta(\alpha, \beta). \tag{70}$$

In fact, $[\lambda_\nu]$ are the eigenvalues and $[\Phi_\nu]$ are the corresponding eigenfunctions of the linear transformation $\mathbf{R}$ on $H(K)$ to itself defined by

$$\mathbf{R}\, h(t) = (h, R(\cdot, t))_{H(K)}. \tag{71}$$

Further

$$\sum_{\nu=1}^{\infty} \lambda_\nu^2 = \|R\|^2_{H(K)\otimes H(K)} < \infty. \tag{72}$$

For $n = 1, 2, \cdots$, let

$$S_n(t) = \sum_{\nu=1}^{n} \eta_\nu \Phi_\nu(t), \quad V_n = (X, \Phi_n)_K. \tag{73}$$

By the developments of Section 1, it follows that P_{S_n+N} is absolutely continuous with respect to P_N with probability density function

$$\begin{aligned} p_n = \frac{dP_{S_n+N}}{dP_N} &= \prod_{\nu=1}^{n} \int_{-\infty}^{\infty} \exp\,(\eta_\nu V_\nu - \tfrac{1}{2}\eta_\nu^2)\, \frac{1}{(2\pi\lambda_\nu)^{1/2}} \exp\left(\frac{-\eta_\nu^2}{2\lambda_\nu}\right) d\eta_\nu \\ &= \prod_{\nu=1}^{n} (1 + \lambda_\nu)^{-1/2} \exp\left(\frac{1}{2}\, V_\nu^2\, \frac{\lambda_\nu}{1 + \lambda_\nu}\right). \end{aligned} \tag{74}$$

By martingale theory it may be shown that (72) implies that the probability density function exists and is given by the limit

$$\frac{dP_{S+N}}{dP_N} = \lim_{n\to\infty} p_n \tag{75}$$

so that

$$\log \frac{dP_{S+N}}{dP_N} = \sum_{\nu=1}^{\infty}\left[-\tfrac{1}{2}\log(1+\lambda_\nu) + \tfrac{1}{2}V_\nu^2 \frac{\lambda_\nu}{1+\lambda_\nu}\right]. \tag{76}$$

If in addition to (72)

$$\sum_{\nu=1}^{\infty} \lambda_\nu < \infty, \tag{77}$$

then the probability density function may be written

$$\log \frac{dP_{S+N}}{dP_N} = -\tfrac{1}{2}\sum_{\nu=1}^{\infty}\log(1+\lambda_\nu) + \tfrac{1}{2}\sum_{\nu=1}^{\infty} V_\nu^2 \frac{\lambda_\nu}{1+\lambda_\nu}. \tag{78}$$

The intuitive meaning of (77) is that almost all sample functions of the signal process $[S(t), t \in T]$ belong to $H(K)$, since from (68)

$$\|S\|_K^2 = \sum_{\nu=1}^{\infty} \eta_\nu^2,$$

$$E[\|S\|_K^2] = \sum_{\nu=1}^{\infty} \lambda_\nu.$$

It appears to establish (77) it would suffice to prove that

$$E[\|S\|_K^2] < \infty.$$

In order to obtain necessary and sufficient conditions that P_{S+N} be absolutely continuous to P_N in the stochastic signal case, let us begin by rephrasing the problem. Let K_1 and K_2 be two positive definite covariance kernels, and let P_i be the probability measure induced on Ω by a normal process $[X(t), t \in T]$ with zero means and covariance kernel K_i. The following questions arise:

1. Determine whether P_1 and P_2 are orthogonal.
2. Determine dP_2/dP_1 if it exists.

We use equations (10) to (15). Let

$$p_{T'} = \frac{dP_{2,T'}}{dP_{1,T'}} = \frac{|K_{2,T'}|^{-1/2}\exp(-\frac{1}{2}X^{\text{tr}}K_{2,T'}^{-1}X)}{|K_{1,T'}|^{-1/2}\exp(-\frac{1}{2}X^{\text{tr}}K_{1,T'}^{-1}X)}, \tag{79}$$

$$\begin{aligned} J_{T'} &= E_{P_2}(\log p_{T'}) - E_{P_1}(\log p_{T'}) \\ &= \tfrac{1}{2}\operatorname{trace}(K_{1,T'}^{-1}K_{2,T'} - I - I + K_{2,T'}^{-1}K_{1,T'}). \end{aligned} \tag{80}$$

Amazingly enough, the right-hand side of (80) can be expressed as the norm of a function in the reproducing kernel Hilbert space corresponding to the kernel $K_1 \otimes K_2$, which is a function of four variables (s, s', t, t') defined by

$$K_1 \otimes K_2(s, s', t, t') = K_1(s, t)\, K_2(s', t'). \tag{81}$$

If K_1 and K_2 are nonsingular covariance matrices, we may verify that

$$\text{trace}\,(K_1 K_2^{-1}) = (K_1, K_1)_{K_1 \otimes K_2}, \tag{82}$$

since

$$\begin{aligned}(K_1, K_1)_{K_1 \otimes K_2} &= \sum_{s,s',t,t'} K_1(s, s')\, K_1^{-1}(s, t)\, K_2^{-1}(s', t')\, K_1(t, t') \\ &= \sum_{s',t,t'} \delta(s', t)\, K_2^{-1}(s', t')\, K_1(t, t') \\ &= \sum_{t,t'} K_2^{-1}(t, t')\, K_1(t, t') \\ &= \text{trace}\,(K_1 K_2^{-1}).\end{aligned} \tag{83}$$

It may also be proved that

$$\text{trace}\, I = (K_1, K_2)_{K_1 \otimes K_2}. \tag{84}$$

In this manner we may verify that

$$\text{trace}\,(K_1 K_2^{-1} + K_2 K_1^{-1} - 2I) = \|K_1 - K_2\|_{K_1 \otimes K_2} \tag{85}$$

$$X^{\text{tr}} K_1^{-1} X - X^{\text{tr}} K_2^{-1} X = (K_2 - K_1, X \otimes X)_{K_1 \otimes K_2}, \tag{86}$$

where $X \otimes X$ is the function on $T \otimes T$ defined by

$$X \otimes X(s, t) = X(s)\, X(t). \tag{87}$$

Using (85) and (86), we may rewrite (79) and (80):

$$p_{T'} = |K_{2,T'}^{-1} K_{1,T'}|^{1/2} \exp\left[\tfrac{1}{2}(K_2 - K_1, X \otimes X)_{K_1 \otimes K_2, T' \otimes T'}\right] \tag{88}$$

$$J_{T'} = \tfrac{1}{2}\|K_2 - K_1\|_{K_1 \otimes K_2, T' \otimes T'}. \tag{89}$$

The following conclusions can be immediately inferred:

1. In order that P_1 and P_2 be orthogonal, it is necessary and sufficient that it is *not* so that

$$K_2 - K_1 \quad \text{belongs to } H(K_1 \otimes K_2;\, T \otimes T). \tag{90}$$

2. If (90) holds, then the Radon-Nikodym derivative exists and is given by the limit (as $T' \to T$) of (88). Formally, we may write

$$\frac{dP_2}{dP_1} = D(K_2^{-1} K_1) \exp\left[\tfrac{1}{2}(K_2 - K_1, X \otimes X)_{K_1 \otimes K_2, T \otimes T}\right] \tag{91}$$

if

$$D(K_2^{-1} K_1) = \lim_{T' \to T} |K_{2,T'}^{-1} K_{1,T'}|^{1/2} \tag{92}$$

is assumed to exist.

By using (91), we can sketch a proof of Woodward's theorem on linear transformation of Wiener integrals [Woodward (1961)].

Example 4. To illustrate the use of (67), we consider stationary time series with spectral density functions, so that

$$R(s-t) = \int_{-\infty}^{\infty} e^{i\omega(s-t)} f_S(\omega)\, d\omega,$$

$$K(s-t) = \int_{-\infty}^{\infty} e^{i\omega(s-t)} f_N(\omega)\, d\omega.$$

We now show that a sufficient condition for (67) to hold for any finite interval $T = (t: 0 \leqslant t \leqslant T)$ is that

$$\int_{-\infty}^{\infty} \frac{f_S(\omega)}{f_N(\omega)}\, d\omega < \infty. \tag{93}$$

To prove (93), we write

$$\begin{aligned}
\|R\|^2_{K\otimes K, T\otimes T} &= \left\| \int_{-\infty}^{\infty} e^{i\omega s} e^{-i\omega t} f_S(\omega)\, d\omega \right\|^2_{K\otimes K, T\otimes T} \\
&= \int_{-\infty}^{\infty} d\omega_1 \int_{-\infty}^{\infty} d\omega_2 f_S(\omega_1) f_S(\omega_2) (e^{i\omega_1(s-t)}, e^{i\omega_2(s-t)})_{K\otimes K, T\otimes T} \\
&= \int_{-\infty}^{\infty} d\omega_1 \int_{-\infty}^{\infty} d\omega_2 f_S(\omega_1) f_S(\omega_2) \left| (e^{i\omega_1 s}, e^{i\omega_2 s})_{K,T} \right|^2 \\
&\leqslant \left[\int_{-\infty}^{\infty} d\omega\, f_S(\omega) \|e^{i\omega s}\|^2_{K,T} \right]^2.
\end{aligned}$$

From (49) we may deduce that

$$\frac{1}{T} \|e^{i\omega s}\|^2_{K,T} \leqslant \int_{-\infty}^{\infty} d\lambda [f_N(\lambda)]^{-1} \frac{1}{T} \left| \frac{1}{2\pi} \int_0^T e^{is(\omega-\lambda)}\, ds \right|^2. \tag{94}$$

As T tends to ∞, the right-hand side of (94) tends to

$$[2\pi f_N(\omega)]^{-1} \tag{95}$$

as a limit in mean with respect to the finite measure on $-\infty < \omega < \infty$ with density function $f_S(\omega)$. The desired conclusion may now be inferred.

It might be noted that by using (94) and (95) we can give simple proofs of various extensions to continuous parameter time series of theorems on the asymptotic efficiency of least-squares estimates of regression coefficients given for discrete parameter time series by Grenander and Rosenblatt (1957).

REFERENCES

Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc.*, **68** 337–404 (1950).

Grenander, U. and M. Rosenblatt. *Statistical Analysis of Stationary Time Series*, New York: Wiley. Stockholm: Almquist & Wiksell, 1957.

Hajek, J. A property of J-divergence of marginal probability distributions. *Czechoslovak Math. J.*, **8** (83) 460–463 (1958a).

Hajek, J. On a property of normal distribution of any stochastic process (in Russian). *Czechoslovak Math. J.*, **8** (83) 610–618, (1958b). (A translation appears in American Mathematical Society Translations in Probability and Statistics, 1961.)

Hajek, J. On linear estimation theory for an infinite number of observations. *Th. Prob. & Applic.*, **6** 182–193 (1961).

Halmos, P. *Measure Theory.* New York: Van Nostrand, 1950.

Parzen, E. Statistical inference on time series by Hilbert space methods, **I.** Department of Statistics, Stanford University, Technical Report No. 23, January 2, 1959.

Parzen, E. Regression analysis of continuous parameter time series. *Proc. Fourth Berkeley Symp. Prob. Math. Statist.*, **I,** University of California Press, 1961.

Parzen, E. An approach to time series analysis, *Ann. Math. Statist.*, **32** 951–989 (1961).

Riesz, F., and B. Sz-Nagy. *Functional Analysis.* London: Blackie, 1956.

Woodward, D. A. A general class of linear transformation of Wiener integrals, *Trans. Am. Math. Soc.*, **100** 459–480 (1961).