Z. Wahrscheinlichkeitstheorie verw. Geb. 26, 309–316 (1973)

The Reproducing Kernel Hilbert Space Structure of the Sample Paths of a Gaussian Process*

Michael F. Driscoll**

1. Introduction

Let the random variables $Y_1, \dots, Y_n$, defined on a probability space $(\Omega, \mathscr{C}, P)$, have mean vector v and nonsingular covariance matrix K. Then

$$P[(Y_1, \dots, Y_n) \in H(K)] = 1 \tag{1}$$

since the reproducing kernel Hilbert space $H(K)$ is simply Euclidean n-space under the inner-product $(a, b) = a' K^{-1} b$.

Parzen [8] mentions that the analogous statement for a random process $Y = \{Y_t : t \in T\}$ on $(\Omega, \mathscr{C}, P)$ is false if T is a (non-finite) separable metric space and the covariance kernel K of Y is continuous on $T \times T$. In fact, he states that under these conditions almost all the sample paths of Y lie outside $H(K)$.

These facts lead one to ask whether there is some other reproducing kernel Hilbert space $H(R)$ of functions on T which contains all the sample paths of Y. In this paper we use a recent result by Kallianpur [4] to show that if Y is a Gaussian process and R is a continuous symmetric positive-definite kernel on $T \times T$, then

$$P[Y \in H(R)] = 0 \quad \text{or} \quad 1. \tag{2}$$

We also derive conditions, involving only the kernels K and R, which characterize the two cases in (2).

One of the hypotheses required for our results is that almost all the sample paths of Y be continuous functions on T. This is a rather restrictive assumption, even for Gaussian processes. For some recent work on sample continuity of Gaussian processes, the reader is referred to Eaves [2], Garcia, Posner, and Rodemich [3], Nisio [6], and Preston [10].

The statement and proof of the main result (Theorem 3) are given in Section 4. We begin in Section 2 with some definitions and notation. The lemmas and theorems which support the main result are given in Section 3. In Section 5 we discuss an application of the main result to signal-noise problems (the details of this application will be given in a later paper).

Throughout the paper, we exclude the case that T is a finite set—as is necessary in view of Eq. (1) and the remarks immediately following it. The results obtained

* Except for Theorem 3, which is now considerably improved, this paper is a portion of the author's doctoral dissertation at the University of Arizona, written under the direction of Professor P.K. Bhattacharya.

** The author dedicates this paper to the memory of his wife Catherine. May she rest in peace.

here hold, with slight and obvious changes, for the case that T is a countable set, but we do not explicitly include this case in what follows.

The analysis in this paper is by reproducing kernel Hilbert space methods. For material on these Hilbert spaces, see Aronszajn [1]. Parzen [7] gives a good introduction to their use in the study of Gaussian processes.

2. Definitions and Notation

Throughout the paper let T be a fixed separable metric space and let T_0 be a fixed countably dense subset of T, say $T_0=\{t_1, t_2, \ldots\}$. Let N denote the positive integers and F the field of real numbers. For each n in N, define $T_n=\{t_1, \ldots, t_n\}$.

A function $R\colon T\times T\to F$ is called a symmetric nonnegative-definite kernel—or simply a *nonnegative kernel*—if $R(s,t)=R(t,s)$ for all $s, t\in T$ and if for each $n\in N$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i R(s_i, s_j)\, a_j \geqq 0 \tag{3}$$

for all $s_1, \ldots, s_n$ in T and all $a_1, \ldots, a_n$ in F. If, in addition, equality holds in (3) only when $a_1=\cdots=a_n=0$, then R is called a *positive kernel.* Note that a function $R\colon T\times T\to F$ is a nonnegative kernel if and only if it is the covariance kernel of some random process indexed by T.

Given a nonnegative kernel R on $T\times T$, we denote by $R_n=[R(t_i, t_j)]=[r_{ij}]$ the matrix obtained by restricting R to $T_n\times T_n$. When R is positive, we denote by $R_n^{-1}=[r^{ij}]$ the inverses of these matrices. The (i,j)-th element of a matrix A will sometimes be denoted $A(i,j)$. The transpose of a matrix A is denoted A'.

A Hilbert space H of functions $h\colon T\to F$ is a *reproducing kernel Hilbert space* (RKHS) if there is a function $S\colon T\times T\to F$ such that

$$S_t(\cdot)=S(\cdot, t)\in H \qquad (t\in T)$$

and

$$(h, S_t)=h(t) \qquad (h\in H,\ t\in T) \tag{4}$$

where $(\cdot,\cdot)$ is the inner-product in H. Then S is a nonnegative kernel and is called the kernel of H, which is then denoted $H=H(S)$. Conversely, every nonnegative kernel S on $T\times T$ determines a unique RKHS $H(S)$ of functions on T. The inner-product and norm of $H(S)$ are denoted $(\cdot,\cdot)_S$ and $\|\cdot\|_S$. The subscript is dropped when no confusion can result. Property (4) is called the *reproducing property.*

A mapping $L\colon G\to H$ between two real Hilbert spaces is called an *operator* if it is linear, continuous, and bounded, and its norm is denoted $\|L\|$.

When $G=H$, we call L *self-adjoint* if $(Lx, y)=(x, Ly)$ for all x, y H and *nonnegative-definite* if $(Lx, x)\geqq 0$ for all $x\in H$. A self-adjoint, nonnegative-definite operator $L\colon H\to H$ on a separable, infinite-dimensional Hilbert space H is called an operator of finite trace if, for some complete orthonormal system (c.o.n.s) $\{e_n\}$ in H, the series

$$\sum_{n=1}^{\infty} (L e_n, e_n) \tag{5}$$

is (absolutely) convergent. It can be shown, see Schatten [12], that if the series (5) converges for some c.o.n.s. then it converges for any c.o.n.s., the value of the sum being independent of the c.o.n.s. chosen.

3. Preliminary Results

The first of the following lemmas follows from Theorems 2E and 5E of Parzen [7] and the second is a special case of Theorem 6E of the same paper.

Lemma 1. *If R is a continuous nonnegative kernel on $T\times T$ then each function in $H(R)$ is continuous on T.*

Lemma 2. *Let R be a continuous positive kernel on $T\times T$ and let $f\colon T\to F$ be a function. Define $(f,f)_{R,n}=f_n' R_n^{-1} f_n$ where $f_n=(f(t_1),\dots,f(t_n))'$. Then $(f,f)_{R,n}\leqq (f,f)_{R,n+1}$ for all n in N and, for f continuous on T, $f\in H(R)$ if and only if*

$$\lim_{n\to\infty}(f,f)_{R,n}<\infty.$$

The next lemma supplies a computational tool needed in the proof of the main result.

Lemma 3. *If R is a continuous positive kernel on $T\times T$, then $H(R)$ is separable and infinite-dimensional. In fact, there is a c.o.n.s. $\{e_n\colon n\in N\}$ in $H(R)$ with $(R_{t_i}, e_j)=0$ for $i<j$. The set of constants $e_{ij}=e_j(t_i)=(e_j, R_{t_i})$ defines a sequence of nonsingular lower-triangular n-by-n matrices $E_n=[e_{ij}]$ such that $R_n=E_nE_n'$ for all n in N. Also, there is a set of constants e^{ij} such that for all n the n-by-n matrix $[e^{ij}]$ is the inverse if E_n. Moreover,*

$$R_{t_i}=\sum_{k=1}^{n} e_{ik}\, e_k \tag{6}$$

and

$$e_i=\sum_{k=1}^{n} e^{ik} R_{t_k} \tag{7}$$

for all i and all $n\geqq i$.

Proof. Since each element of $H(R)$ is continuous on T (by Lemma 1), $\{R_t\colon t\in T_0\}$ is a spanning set for $H(R)$. Since each matrix R_n is nonsingular, the elements of $\{R_t\colon t\in T_0\}$ are linearly independent. Thus $H(R)$ is infinite-dimensional and hence separable. By Gram-Schmidt orthogonalization there is a c.o.n.s. $\{e_n\}$ in $H(R)$ with $(R_{t_i}, e_j)=0$ for $i<j$. The existence of matrices E_n with the stated properties follows from the linear independence of the e_n's, Parseval's inequality, and the reproducing property of $H(R)$, and the lower triangularity of each E_n guarantees that $E_n^{-1}(i,j)=E_{n+1}^{-1}(i,j)$ for all $n\geqq i,j$. Since $\{R_t\colon t\in T_0\}$ spans $H(R)$ and

$$\begin{aligned}(R_{t_j}, R_{t_i}) &= R(t_i, t_j)=R_n(i,j)\\ &=(E_nE_n')(i,j)=\sum_{k=1}^{n} e_{ik}(e_k, R_{t_j})\\ &=\left(R_{t_j}, \sum_{k=1}^{n} e_{ik}\, e_k\right)\end{aligned}$$

for all $n\geqq i,j$, (6) holds. The proof of (7) is similar.

The proof of the main result also requires knowledge of the set-theoretic relationships between RKHS's. The necessary information is given by the next theorem.

Theorem 1. *Let K be a continuous nonnegative kernel on $T\times T$ and let R be a continuous positive kernel on $T\times T$. Then the following statements are equivalent:*

(a) $H(K)\subset H(R)$, *in which case* (a′) *there is a constant $B>0$ such that $\|g\|_R\leqq B\|g\|_K$ for all $g\in H(K)$.*

(b) *There is $B<\infty$ such that B^2R-K is a nonnegative kernel.*

(c) *There is an operator L: $H(R)\to H(K)$ with $\|L\|\leqq B$ and $LR_t=K_t$ for all $t\in T_0$, in which case* (c′) *$LR_t=K_t$ for all $t\in T$.*

Moreover, each of these implies

(d) *There is a self-adjoint nonnegative-definite operator L: $H(R)\to H(R)$ with $\|L\|\leqq B^2$ and $LR_t=K_t$ for all $t\in T$.*

The constant B is fixed throughout.

Proof. Aronszajn ([1], 382–383) proves that (a)$\Leftrightarrow$(b) and that (a)$\Rightarrow$(a′).

Proof that (b)$\Leftrightarrow$(c). By the proof of Lemma 3, the linear hull H of $\{R_t\colon t\in T_0\}$ is dense in $H(R)$ so L: $H\to H(K)$ can be defined by $LR_t=K_t$ $(t\in T_0)$ and by linearity. If (b) holds then, since each matrix $B^2R_n-K_n$ is nonnegative-definite, L is bounded on H, hence extends uniquely to an operator L: $H(R)\to H(K)$. Conversely, (c) implies $B^2\|h\|_R^2\geqq\|Lh\|_K^2$ for all $h\in H\subset H(R)$ so each matrix $B^2R_n-K_n$ is nonnegative-definite. By the continuity of K and R, B^2R-K is a nonnegative kernel.

Proof that (c)$\Rightarrow$(c′). This follows easily from the facts that L is continuous, that $H(K)$ is Hausdorff, and that $s_n\to s$ in T implies $R_{s_n}\to R_s$ strongly in $H(R)$ and $K_{s_n}\to K_s$ strongly in $H(K)$.

Proof that (a)–(c)$\Rightarrow$(d). If L is the operator in (c) then by (a′)

$$\|Lh\|_R\leqq B\|Lh\|_K\leqq B^2\|h\|_R \qquad (h\in H(R))$$

so that L: $H(R)\to H(R)$ is an operator with $LR_t=K_t$ $(t\in T)$. Thus

$$\begin{aligned}(LR_s, R_t)_R=(K_s, R_t)_R=K(s,t)=K(t,s)=(K_t, R_s)_R\\ =(R_s, LR_t)_R\end{aligned}$$

for all s,t in T. Since $\{R_t\colon t\in T\}$ spans $H(R)$ and L, K, and R are continuous, L is self-adjoint. That L is nonnegative-definite follows similarly from the continuity of L and the fact that K is a nonnegative kernel. The proof of Theorem 1 is complete.

The following theorem is from an example of Prohorov ([11], Sect. 1.6).

Theorem 2. *Let H be an infinite-dimensional separable Hilbert space and let $\mathscr{E}$ be the Borel σ-field from H, that is, the smallest σ-field of subsets of H with respect to which every continuous linear functional f: $H\to F$ is measurable. On $(H,\mathscr{E})$ define the random process $Y^*=\{Y_g^*\colon g\in H\}$ by $Y_g^*(h)=(h,g)_H$ for all $g, h\in H$. Then there is a probability P^* on $(H,\mathscr{E})$ such that Y^* is a Gaussian process under P^* with mean value function m^*: $H\to F$ and covariance kernel K^*: $H\times H\to F$ if and only if there are $m\in H$ and a self-adjoint nonnegative-definite operator L: $H\to H$ of finite trace such that $m^*(g)=(m,g)_H$ and $K^*(g,g')=(Lg,g')_H$ for all $g, g'\in H$.*

4. The Main Result

Let $\mathscr{Y}$ be the set of all continuous functions $y: T \to F$ and let $\mathscr{D}$ be the smallest σ-field from $\mathscr{Y}$ containing all sets of the form $\{y: y(t) \in B\}$ where $t \in T$ and B is a Borel subset of F. If R is a continuous nonnegative kernel on $T \times T$, then $H(R) \subset \mathscr{Y}$ by Lemma 1. If R is also a positive kernel, then Lemma 2 gives

$$H(R) = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \{h \in \mathscr{Y}: (h,h)_n \leqq k\}$$

$$= \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \{h \in \mathscr{Y}: (h(t_1), \ldots, h(t_n)) \in q_n^{-1}([a,b])\}$$

where q_n is the positive-definite quadratic form $q(a_1, \ldots, a_n) = a' R_n^{-1} a$. Thus $H(R)$ is a $\mathscr{D}$-measurable subset of $\mathscr{Y}$. Further, the σ-field $\mathscr{E} = \{D \cap H(R): D \in \mathscr{D}\}$ is the Borel σ-field from $H(R)$, as is easily shown using the reproducing property of $H(R)$.

Having settled these details, we can now prove our main result.

Theorem 3. *Let the random process $Y = \{Y_t: t \in T\}$ on a probability space $(\Omega, \mathscr{C}, P)$ be Gaussian with mean value function v and covariance kernel K. Assume that K is continuous on $T \times T$ and that almost all the sample paths of Y are continuous on T. Let R be a continuous positive kernel on $T \times T$ with $v \in H(R)$. Then either*

$$P[Y \in H(R)] = 1 \quad \text{or} \quad P[Y \in H(R)] = 0. \tag{8}$$

Further, this probability is 1 or 0 according as

$$\tau = \sup_{n \in N} \operatorname{tr}(K_n R_n^{-1}) \tag{9}$$

is finite or infinite, where $\operatorname{tr}(\cdot)$ *is the trace function for finite matrices.*

Proof. Since $v \in H(R)$ and $H(R)$ is a linear space, we assume without loss of generality that $v = 0$. Let Q be the probability measure induced on $(\mathscr{Y}, \mathscr{D})$ by the random process Y, that is, $Q(D) = P(Y \in D)$ for all $D \in \mathscr{D}$. Then it follows immediately from Theorem 1 of Kallianpur [4] that $Q[H(R)] = 0$ or 1, which proves (8).

By Lemma 2 and the Monotone Convergence Theorem we have

$$E_p[\|Y - v\|_R^2] = \lim_{n \to \infty} E_p[\|Y - v\|_{R,n}^2]$$

$$= \lim_{n \to \infty} \operatorname{tr}(K_n R_n^{-1}) = \tau.$$

So if $\tau < \infty$, then $P[\|Y - v\|_R^2 < \infty] = 1$, thus $P[Y \in H(R)] = 1$.

Conversely, assume that $P[Y \in H(R)] = 1$. Let the random process

$$Y_T^* = \{Y_{R_t}^*: t \in T\}$$

and the probability P^* on $(H(R), \mathscr{E})$ be defined by

$$Y_{R_t}^*(h) = (h, R_t)_R = h(t) \quad (h \in H(R), t \in T)$$

and

$$P^*(E) = Q(E) \quad (E \in \mathscr{E}).$$

Since $P^*[Y_T^* \in E] = P^*(E) = Q(E) = P(Y \in E)$ for all E in $\mathscr{E}$, Y_T^* is a Gaussian process under P^* with mean value function zero and covariance kernel K.

By Lemma 5 of Kallianpur [4], $P[Y \in H(R)] = 1$ implies $H(K) \subset H(R)$, so by Theorem 1 above there is a self-adjoint nonnegative-definite operator L: $H(R) \to H(R)$ such that $LR_t = K_t \, (t \in T)$. Define the nonnegative kernel K^*: $H(R) \times H(R) \to F$ by

$$K^*(g, g') = (Lg, g')_R \qquad (g, g' \in H(R)). \tag{10}$$

Then $K^*(R_s, R_t) = K(s, t) \, (s, t \in T)$ so Y_T^* under P^* is Gaussian with mean zero and covariance kernel K^* on $\{R_t: t \in T\} \times \{R_t: t \in T\}$.

Let H be the linear hull of $\{R_t: t \in T\}$ and let $Y_H^* = \{Y_g^*: g \in H\}$ be the random process defined by $Y_g^*(h) = (h, g)_R$ for all h in $H(R)$ and g in H. Then clearly Y_H^* under P^* is Gaussian with mean zero and covariance kernel K^*: $H \times H \to F$.

Finally, consider the random process $Y^* = \{Y_g^*: g \in H(R)\}$ on $(H(R), \mathscr{E})$ defined by $Y_g^*(h) = (h, g)_R \, (g, h \in H(R))$. It then follows from the facts that H is dense in $H(R)$, that $g_n \to g$ strongly in $H(R)$ implies $Y_{g_n}^* \to Y_g^*$ a.s. (P^*), and the Dominated Convergence Theorem that Y^* under P^* is Gaussian with mean zero and covariance kernel K^*.

So by Theorem 2, there is a self-adjoint nonnegative-definite operator

$$L^*: \; H(R) \to H(R)$$

of finite trace such that

$$K^*(g, g') = (L^* g, g')_R \qquad (g, g' \in H(R)).$$

In view of (10), $L^* = L$, so L is an operator of finite trace. Therefore

$$\sum_{i=1}^{\infty} (Le_i, e_i)_R < \infty$$

where $\{e_n\}$ is the c.o.n.s. of Lemma 3. But by (7)

$$\begin{aligned}
\sum_{i=1}^{n} (Le_i, e_i)_R &= \sum_{i=1}^{n} \sum_{p=1}^{n} \sum_{q=1}^{n} e^{ip} (LR_{t_p}, R_{t_q})_R \, e^{iq} \\
&= \sum_{i=1}^{n} \sum_{p=1}^{n} \sum_{q=1}^{n} e^{ip} K_n(p, q) \, e^{iq} \\
&= \operatorname{tr}[E_n^{-1} K_n (E_n^{-1})'] \\
&= \operatorname{tr}[K_n R_n^{-1}],
\end{aligned}$$

so $\tau < \infty$. This concludes the proof.

It was noted in the proof of Theorem 3 that the supremum in (9) may be replaced by limit. It is also true that $\tau < \infty$ if and only if the sequence $\{d_n\}$ where

$$d_n = |R_n (R_n + K_n)^{-1}| \qquad (n \in N)$$

and $|A|$ denotes the determinant of the matrix A, has a finite nonzero limit d (in which case $0 < d \leqq 1$). The proof of this equivalence requires a theorem on infinite matrices in MacDuffee ([5], 105) and properties of the matrices $K_n R_n^{-1}$ which follow from Lemma 3. The details of the proof are mainly computational and are therefore omitted.

5. An Application

The result of the previous section has an application to the following *signal-noise problem.*

Suppose that $S=\{S_t\colon t\in T\}$ and $N=\{N_t\colon t\in T\}$ are random processes on a probability space $(\Omega, \mathscr{C}, P)$ such that, under P, S is Gaussian with mean zero and covariance kernel K, N is Gaussian with mean zero and covariance kernel R, and S and N are independent. Given a sample path $x=s+n$ from the random process $X=S+N$, the problem is to estimate the corresponding sample path s from S, that is, to eliminate the *noise* n from the data x to recover the *signal* s.

In order to obtain such an estimate based on x, it is first necessary to determine which signals s could have produced x and then focus one's attention on that class of signals. This is precisely where Theorem 3 applies. For if K and R are such that τ of (9) is finite, then $P[S\in H(R)]=1$ and so, by a result of E. Parzen, the conditional distributions of X given $S=s$ are all mutually absolutely continuous—that is, the class of signals to be considered (for any x) is the class of all possible signals. Further, the class of all possible signals has Hilbert space structure.

In this framework, formulas can be derived for the conditional probability density functionals of X given $S=s$ and of S given $X=x$ and for the marginal probability density functionals of S and X. It can be shown that under the generalized square-error loss function

$$L(a, b)=\|a-b\|^2,$$

where $\|\cdot\|$ is the norm in a RKHS, the best estimate of s given x is the conditional mean of S given $X=x$ and also that this estimate is the limit on T of a sequence of functions which have closed forms involving x, K, and R. A similar approximating sequence can be found for the conditional expected loss associated with this estimate.

The proofs of these facts employ several results of Parzen [9] about Gaussian processes. The details of these arguments will be given in a later paper.

References

1. Aronszajn, N.: Theory of reproducing kernels. Trans. Amer. Math. Soc. **68**, 337–404 (1950).
2. Eaves, D. M.: Sample functions of Gaussian random fields are either continuous or very irregular. Ann. Math. Statist. **38**, 1579–1582 (1967).
3. Garcia, A. M., Posner, E. C., Rodemich, R. R.: Some properties of the measures on function spaces induced by Gaussian processes. J. Math. Anal. Appl. **21**, 150–161 (1968).
4. Kallianpur, G.: Zero-one laws for Gaussian processes. Trans. Amer. Math. Soc. **149**, 199–211 (1970).
5. MacDuffee, C. C.: The Theory of Matrices. New York: Chelsea 1946.
6. Nisio, M.: On the continuity of stationary Gaussian processes. Nagoya Math. J. **34**, 89–104 (1969).
7. Parzen, E.: Statistical Inference on Time Series by Hilbert Space Methods, I. Dept. of Statist., Stanford Univ., Technical Report No. 23, January 2, 1959.
8. Parzen, E.: Probability density functionals and reproducing kernel Hilbert spaces. Time Series Analysis. 155–169 M. Rosenblatt, Ed. New York: Wiley 1963.

9. Parzen, E.: Time Series Analysis Papers. San Francisco: Holden-Day 1967.
10. Preston, C.: Continuity properties of some Gaussian processes. Ann. Math. Statist. **43**, 285–292 (1972).
11. Prohorov, Yu. V.: Convergence of random processes and limit theorems in probability theory. Theor. Probab. Appl. **1**, 157–214 (1956).
12. Schatten, R.: A Theory of Cross Spaces. Ann. Math. Studies **26**, 1–153 (1950).

Michael F. Driscoll
Department of Mathematics
Arizona State University
Tempe, Arizona 85281, USA

(Received August 30, 1972)