TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 353, Number 10, Pages 3945–3969
S 0002-9947(01)02852-5
Article electronically published on May 14, 2001

STOCHASTIC PROCESSES WITH SAMPLE PATHS IN REPRODUCING KERNEL HILBERT SPACES

MILAN N. LUKIĆ AND JAY H. BEDER

Abstract. A theorem of M. F. Driscoll says that, under certain restrictions, the probability that a given Gaussian process has its sample paths almost surely in a given reproducing kernel Hilbert space (RKHS) is either 0 or 1. Driscoll also found a necessary and sufficient condition for that probability to be 1.

Doing away with Driscoll's restrictions, R. Fortet generalized his condition and named it *nuclear dominance*. He stated a theorem claiming nuclear dominance to be necessary and sufficient for the *existence* of a process (not necessarily Gaussian) having its sample paths in a given RKHS. This theorem – specifically the necessity of the condition – turns out to be incorrect, as we will show via counterexamples. On the other hand, a weaker sufficient condition is available.

Using Fortet's tools along with some new ones, we correct Fortet's theorem and then find the generalization of Driscoll's result. The key idea is that of a random element in a RKHS whose values are sample paths of a stochastic process. As in Fortet's work, we make almost no assumptions about the reproducing kernels we use, and we demonstrate the extent to which one may dispense with the Gaussian assumption.

1. Introduction

Consider a Gaussian process $\{X_t, t \in T\}$ on a probability space $\{\Omega, \mathcal{A}, \mathbf{P}\}$ having mean and covariance functions m and K, and let R be a positive symmetric kernel on $T \times T$. Let $X_{\bullet}$ denote the sample path of X, and let $\mathcal{H}(R)$ denote the reproducing kernel Hilbert space with kernel R. Driscoll [3, Theorem 3] proved that, under appropriate assumptions,

$$\mathbf{P}(X_{\bullet} \in \mathcal{H}(R)) = 0 \text{ or } 1,$$

and that the probability is 1 iff a certain limit τ is finite. We will refer to this result as *Driscoll's Theorem.* Driscoll applied this in [4] to find the Bayes estimator of m under a Gaussian prior, using the norm of $\mathcal{H}(R)$ to define a quadratic loss function.

Driscoll's assumptions are

1. that T is a separable metric space, that K and R are continuous on $T \times T$, and that R is positive definite; and
2. that the sample paths of X are continuous,

Received by the editors March 10, 2000 and, in revised form, February 8, 2001.
2000 *Mathematics Subject Classification.* Primary 60G12, Secondary 60B11, 60G15, 28C20, 46E22, 47B32.
Key words and phrases. Covariance operator, Gaussian process, nuclear dominance, random element in Hilbert space, reproducing kernel Hilbert space, second order process, zero-one law.

an assumption which Driscoll calls "rather restrictive". These assumptions allow him to assert the measurability of $\mathcal{H}(R)$ and thus to apply the zero-one law of Kallianpur [9]. To define the number τ, Driscoll fixes a countable dense subset $T_0 = \{t_1, t_2, \dots\}$ of T and defines the matrices K_n and R_n to be the restrictions of the kernels K and R to $\{t_1, \dots, t_n\}$. Assuming that R is positive definite, τ is then defined by

$$\tau = \sup_n \operatorname{Tr}\left(K_n R_n^{-1}\right). \tag{1.1}$$

(In [2] Driscoll also proves his result when T is a countable set, in which case the measurability of $\mathcal{H}(R)$ is easy to show.)

Two key steps in generalizing Driscoll's Theorem are (1) the recognition that τ is the trace of an operator (which we will call the *dominance operator*) on $\mathcal{H}(R)$, and (2) the definition of an intrinsic pseudo-metric d_R on the set T and of a *Hamel subset* T_0 of T (Section 4). The second of these requires no assumptions on R or K or on the set T, while the first, due to Fortet [6], requires only that R *dominate* K (his terminology) in the sense that $\mathcal{H}(R) \supset \mathcal{H}(K)$. The finiteness of τ Fortet naturally calls *nuclear dominance*. We will use $R \geq K$ to denote dominance, and $R \gg K$ if the dominance is nuclear.

Thus we may be guided by the following:

Working Conjecture. *Let X be a process with covariance K. In order that $\mathbf{P}(X_{\cdot} \in \mathcal{H}(R)) = 1$, it is necessary and sufficient that $R \gg K$.*

In the Gaussian case this turns out to be true, and moreover we are able to extend Driscoll's zero-one law to this setting (Theorems 7.4 and 7.5). In the general case, however, something different emerges. Here we must take necessity and sufficiency separately.

Necessity turns out to be false, contrary to [6, Theorem 2], as the counterexamples of Section 2 show. On the other hand, sufficiency is correct (in a sense we make precise in Theorem 5.1) but requires some significant additional machinery to cope with the lack of a metric on T and with the possibility that R is semi-definite. We develop this machinery in Section 4. Theorem 5.1 provides the strongest answer we are able to achieve regarding the sufficiency of nuclear dominance in the Working Conjecture.

In [6, Theorem 2] Fortet actually makes a slightly different assertion from our Working Conjecture, in the form of an existence result: *Let K and R be two reproducing kernels such that $\mathcal{H}(K)$ and $\mathcal{H}(R)$ are separable.*

In order that there exist a second-order random process $\{X_t,\, t \in T\}$ with covariance K and with trajectories in $\mathcal{H}(R)$ with probability 1*, it is necessary and sufficient that $R \gg K$.*

Moreover, when $R \gg K$, there exists a zero-mean Gaussian process $\{X_t, t \in T\}$ with trajectories in $\mathcal{H}(R)$ almost surely.

Fortet published this result without proof. The necessity assertion is, of course, identical to that of our Working Conjecture, since assuming the existence of a process is the same as being given a process X. We have already pointed out that this assertion is true in the Gaussian case but not in general. On the other hand, the sufficiency assertion is weaker than Driscoll's, since proving the existence of a process with a certain property is weaker than proving that a given process has that property. It may not be surprising therefore that a weaker condition, namely

$R \geq K$, is sufficient to guarantee the existence of a *non*-Gaussian process with covariance K and almost all sample paths in $\mathcal{H}(R)$ (Theorem 6.1). The existence of a *Gaussian* process, however, requires the stronger assumption of nuclear dominance (Theorem 7.2).

A cornerstone of our analysis is the relation between a random process X (with sample paths in $\mathcal{H}(R)$) and the random element ξ in $\mathcal{H}(R)$ which it defines, and particularly the role played by the measure induced on $\mathcal{H}(R)$. The second order properties of ξ, which are different from (and stronger than) that of X, turn out to be exactly what we need. We discuss the random element approach in Section 2. First we need to make precise some of the concepts we have discussed in this introduction.

1.1. **Notation and background.** The set of real numbers will be denoted by $\mathbb{R}$, and the natural numbers by $\mathbb{N}$. All Hilbert spaces we consider are over $\mathbb{R}$.

A (real-valued) *random process* on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ is a family $X = \{X_t,\, t \in T\}$ of random variables on Ω with values in $\mathbb{R}$. That is, for each t, X_t is a Borel-measurable function on Ω. We denote the *trajectory* or *sample path* of X at ω by $X_{\bullet}(\omega)$. Thus $X_{\bullet} \in B$ means that the trajectory of X belongs to the set B, so that $\{X_{\bullet} \in B\}$ corresponds to the set $\{\omega \in \Omega : X_{\bullet}(\omega) \in B\}$. We say that the process Y is a *version* of X if $\mathbf{P}(X_t = Y_t) = 1$ for every $t \in T$.

The process X has *p-th order* if $\mathbf{E}|X_t|^p < \infty$ for each $t \in T$. A second-order process X has *mean function* m defined by $m(t) = \mathbf{E}(X_t)$ and *covariance function* K defined by $K(s,t) = \mathbf{E}\left[(X_s - m(s))\,(X_t - m(t))\right]$.

Let K be the covariance function of the process X. Then K is symmetric, and is *positive* in the sense that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j K(t_i, t_j) \geq 0$$

for all n, all $t_1, \dots t_n \in T$, and $a_1, \dots, a_n \in \mathbb{R}$. A positive symmetric function on $T \times T$ will be referred to as a *covariance kernel* or a *reproducing kernel*. By slight abuse of terminology, we will refer to K as a kernel *on* T. The kernel K is *positive definite* or *nonsingular* if the inequality above is always strict.

A covariance kernel K gives rise to a unique Hilbert space $\mathcal{H}(K)$ whose elements are real functions on T, and such that

$$\begin{aligned} &K_t \equiv K(t,\cdot) \in \mathcal{H}(K) \quad \text{for all } t \in T, \text{ and} \\ &\langle f, K_t \rangle = f(t) \quad \text{for all } f \in \mathcal{H}(K), \text{ and all } t \in T, \end{aligned} \tag{1.2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathcal{H}(K)$. The Hilbert space $\mathcal{H}(K)$ is called *a reproducing kernel Hilbert space* (RKHS) with the reproducing kernel K [1]. The property (1.2) is referred to as the *reproducing property*.

If we need to indicate the index set T, we will write $\mathcal{H}(K,T)$ for $\mathcal{H}(K)$. Similarly, if we need to specify the inner product or norm in $\mathcal{H}(K)$ or $\mathcal{H}(R)$, we will use the notation $\langle \cdot, \cdot \rangle_K$, $\|\cdot\|_R$, etc. Further related notation will be introduced in Section 4.

The following lemma illustrates some of the importance of RKHS in the study of second-order random processes. Let $\{X_t,\, t \in T\}$ be a second-order process, defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. The smallest Hilbert subspace of $L^2(\Omega, \mathcal{A}, \mathbf{P})$ that contains all the variables X_t, $t \in T$, is called *the Hilbert space spanned by the process* X, or simply *the Hilbert space of* X.

Lemma 1.1 (Loève [11])**.** *Let H be the Hilbert space of a zero-mean second-order process X with covariance K. Then H is congruent to the RKHS $\mathcal{H}(K)$. The corresponding isometry $\Lambda : H \to \mathcal{H}(K)$ is given by*

$$\Lambda(Y)(t) = \mathbf{E}(YX_t), \quad t \in T. \tag{1.3}$$

The map Λ defined by (1.3) is referred to as *Loève's isometry*.

1.2. **Dominance.** We say that R *dominates* K, and write $R \geq K$, or $K \leq R$, if $\mathcal{H}(K) \subseteq \mathcal{H}(R)$ [5]. Note that $\mathcal{H}(K)$ is a vector subspace of $\mathcal{H}(R)$ but in general has a different inner product and so is not a Hilbert subspace of $\mathcal{H}(R)$.

Theorem 1.1. *Let $R \geq K$. Then*

$$\|g\|_R \leq \|g\|_K, \quad \forall g \in \mathcal{H}(K). \tag{1.4}$$

Moreover, there exists a unique linear operator $L : \mathcal{H}(R) \to \mathcal{H}(R)$ whose range is contained in $\mathcal{H}(K)$, and such that

$$\langle f, g\rangle_R = \langle Lf, g\rangle_K, \quad \forall f \in \mathcal{H}(R), \forall g \in \mathcal{H}(K). \tag{1.5}$$

In particular,

$$LR_t = K_t, \quad \text{for all } t \in T. \tag{1.6}$$

As an operator into $\mathcal{H}(R)$, L is bounded, positive and symmetric.

Conversely, let $L : \mathcal{H}(R) \to \mathcal{H}(R)$ be a positive, continuous, selfadjoint operator. Then $K(s,t) = \langle LR_s, R_t\rangle_R$, $s, t \in T$, defines a reproducing kernel on T such that $K \leq R$.

This theorem gathers together a set of results found in [1, pages 354f, 372f, and 382f]. Part of it is stated as Theorem 1.4 in [5].

We will call the operator L in the theorem the *dominance operator of* $\mathcal{H}(R)$ *over* $\mathcal{H}(K)$. We say that the dominance is *nuclear* if L is nuclear [6], in which case we write $R \gg K$.

2. Random elements in RKHS

Our main tool in the treatment of a process X with sample paths in a RKHS $\mathcal{H}(R)$ is the random element ξ whose realizations in $\mathcal{H}(R)$ are the sample paths of X. This section gives a precise description of that relationship.

For a Hilbert space H, let $\hat{\mathfrak{C}}(H)$ denote the cylindrical σ-algebra on H, and for a topological space $\mathcal{X}$ let $\mathfrak{B}(\mathcal{X})$ be the Borel σ-algebra on $\mathcal{X}$. When $\mathcal{X}$ is a Hilbert space H, $\mathfrak{B}(H)$ is defined by the norm topology on H, and we have $\hat{\mathfrak{C}}(H) \subseteq \mathfrak{B}(H)$.

Finally, we say that a subset D of a topological space $\mathcal{X}$ is *separable* if there exists a countable subset $E \subset \mathcal{X}$ such that $D \subseteq \overline{E}$, where $\overline{E}$ denotes the closure of E. Of course, if D is a separable set, so is $\overline{D}$ and so is any subset of D.

Definition 2.1. Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space and let $(\Omega', \mathcal{A}')$ be a measure space. A *random element in* $(\Omega', \mathcal{A}')$ is an $(\mathcal{A}', \mathcal{A})$-measurable function $\xi : \Omega \to \Omega'$, that is, a function such that $\xi^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{A}'$.

If Ω' is a Hilbert space H and $\mathcal{A}' = \hat{\mathfrak{C}}(H)$, then ξ is called a *random element in* H.

If Ω' is a topological space, then ξ is *separably valued* if $\xi(\Omega)$ is a separable subset of Ω', and is *Borel* if $\mathcal{A}' = \mathfrak{B}(\Omega')$.

It follows from the definition of the cylindrical σ-algebra $\hat{\mathfrak{C}}(H)$ that a map $\eta : \Omega \to H$ is a random element in H iff $\langle \eta, h\rangle$ is a random variable for every $h \in H$.

Lemma 2.1. *Let $\mathcal{H}(R)$ be a RKHS and let X be a stochastic process on $(\Omega, \mathcal{A}, \mathbf{P})$ with almost all sample paths in $\mathcal{H}(R)$. Then*

$$\xi(\omega) = X_{\bullet}(\omega), \quad \omega \in \Omega, \tag{2.1}$$

defines a random element in $\mathcal{H}(R)$, and we have

$$X_t = \langle \xi, R_t\rangle_R, \quad t \in T. \tag{2.2}$$

Conversely, let ξ be a random element in $\mathcal{H}(R)$ defined on $(\Omega, \mathcal{A}, \mathbf{P})$. Then (2.2) *defines a stochastic process on $(\Omega, \mathcal{A}, \mathbf{P})$, and* (2.1) *holds.*

Every separably valued random element in $\mathcal{H}(R)$ is a Borel random element in $\mathcal{H}(R)$.

Proof. Assume first that X is a random process with sample paths in $\mathcal{H}(R)$, and define ξ by (2.1). Let V be the linear span of $\{R_t, t \in T\}$. An arbitrary $h \in V$ has the form $h = \sum_{k=1}^n a_k R_{t_k}$, where $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$, $t_1, \dots, t_n \in T$. Thus $\langle \xi, h\rangle = \sum_{k=1}^n a_k X_{t_k}$ is a random variable for every $h \in V$. For an arbitrary $h \in \mathcal{H}(R)$, Lemma 4.1 below asserts the existence of a sequence (h_n) in V such that $\lim \|h_n - h\| = 0$. Hence,

$$\lim_{n\to\infty} \langle \xi, h_n\rangle = \langle \xi, h\rangle,$$

for all $\omega \in \Omega$, i.e., $\langle \xi, h\rangle$ is a random variable, as a pointwise limit of the sequence $(\langle \xi, h_n\rangle)$ of random variables. Hence, ξ is a random element in $\mathcal{H}(R)$.

On the other hand, if ξ is a random element in $\mathcal{H}(R)$, then $\langle \xi, h\rangle$ defines a random variable for every $h \in \mathcal{H}(R)$. In particular, each X_t defined by (2.2) is a random variable, i.e., X is a random process with sample paths in $\mathcal{H}(R)$.

The fact that a separably valued random element in a Hilbert space is necessarily Borel is contained in [16, Theorem II.1.1]. □

Let ξ be a random element in a measurable space $(\mathcal{X}, \mathfrak{B})$. The measure $\mathbf{P}_\xi$ given by

$$\mathbf{P}_\xi(B) = \mathbf{P}\big(\xi^{-1}(B)\big), \quad B \in \mathfrak{B},$$

is called *the (probability) distribution* of the random element ξ. Let $\mathcal{X}$ be a Banach space and $0 < p < \infty$. A measure μ on $\hat{\mathfrak{C}}(\mathcal{X})$ is said to have *weak p-th order* if

$$\int_{\mathcal{X}} |\langle y, x^\star\rangle|^p d\mu(y) < \infty$$

for all $x^\star \in \mathcal{X}^\star$.

A *Borel* measure μ on $\mathcal{X}$ is said to have *strong p-th order* if

$$\int_{\mathcal{X}} \|x\|^p d\mu(x) < \infty.$$

We require μ to be Borel here so that $\|x\|$ be measurable.

A random element ξ has *weak (strong) order p* if its distribution $\mathbf{P}_\xi$ has the corresponding property.

The concept of mathematical expectation extends to random elements via the notion of the Pettis integral.

Theorem 2.1. *Let $(\Omega, \mathfrak{A}, \mathbf{P})$ be a probability space, $\mathcal{H}$ a Hilbert space, and $\xi : \Omega \to \mathcal{H}$ a random element with weak first order. Then there exists an element $\mathbf{E}\xi \in \mathcal{H}$ such that*

$$\int_\Omega \langle \xi(\omega), h\rangle d\mathbf{P}_\xi(\omega) = \langle \mathbf{E}\xi, h\rangle \tag{2.3}$$

holds for all $h \in \mathcal{H}$.

See [16, Proposition II.3.1]. Equation (2.3) may be written as

$$\mathbf{E}\langle \xi, h\rangle = \langle \mathbf{E}\xi, h\rangle.$$

The element $\mathbf{E}\xi$ defined in Theorem 2.1 is called *the Pettis integral*, or *mathematical expectation*, or *mean* of the random element ξ.

Theorem 2.2. *Let $\mathcal{H}$ be a Hilbert space, and let μ be a weak second order probability measure defined on the cylindrical σ-algebra $\hat{\mathfrak{C}}(\mathcal{H})$. The bilinear form*

$$r_\mu(f, g) = \int_\mathcal{H} \langle x, f\rangle \langle x, g\rangle d\mu(x) - \int_\mathcal{H} \langle x, f\rangle\, d\mu(x) \int_\mathcal{H} \langle x, g\rangle\, d\mu(x)$$

is defined for all $f, g \in \mathcal{H}$, and is continuous with respect to the norm topology of $\mathcal{H}$.

The operator Θ defined by

$$\langle \Theta f, g\rangle = r_\mu(f, g) \tag{2.4}$$

is a continuous symmetric linear operator in $\mathcal{H}$.

The operator Θ defined in (2.4) is referred to as the *covariance operator* of the measure μ. The covariance operator of a (necessarily weak second order) random element ξ is the covariance operator of its distribution $\mathbf{P}_\xi$. Thus the defining equation for the covariance operator of a random element ξ in a Hilbert space $\mathcal{H}$ is

$$\langle \Theta f, g\rangle = \mathbf{E}\langle \xi, f\rangle \langle \xi, g\rangle - \langle \mathbf{E}\xi, f\rangle \langle \mathbf{E}\xi, g\rangle, \quad f, g \in \mathcal{H}. \tag{2.5}$$

Theorem 2.2 and the subsequent definitions are based on [16, Sections III.1 and III.2]. A brief but rather complete account can also be found in [23, pages 11f].

When $\mathcal{X}$ is a RKHS, we may connect the order of ξ with the order of the process it defines:

Theorem 2.3. *Let ξ be a random element in $\mathcal{H}(R)$, and let X be the process it defines.*

1. *If ξ has weak first order, then X has first order, and*

$$\mathbf{E}(X_t) = \langle \mathbf{E}\xi, R_t\rangle_R, \quad t \in T.$$

In particular, the mean function m of X belongs to $\mathcal{H}(R)$, and $m = \mathbf{E}\xi$.

2. *If ξ has weak second order, then X has second order. In this case, if Θ is the covariance operator of ξ and K is the covariance function of X, then*

$$K(s, t) = \langle \Theta R_s, R_t\rangle_R. \tag{2.6}$$

In particular, $R \geq K$ and Θ is the dominance operator.

Proof. (1) $\mathbf{E}\xi \in \mathcal{H}(R)$ exists by Theorem 2.1. Each X_t is integrable because

$$\mathbf{E}|X_t| = \mathbf{E}|\langle \xi, R_t\rangle| < \infty,$$

by the assumption of weak first order of ξ. Moreover, using (2.3),

$$m(t) \equiv \mathbf{E}[X_t] = \mathbf{E}\langle \xi, R_t \rangle = \langle \mathbf{E}\xi, R_t \rangle = (\mathbf{E}\xi)(t),$$

for all $t \in T$.

(2) The fact that X is a second-order process follows from

$$\mathbf{E}X_t^2 = \mathbf{E}\langle \xi, R_t \rangle_R^2 < \infty,$$

by the assumption of weak second order. To prove (2.6), note first that ξ has weak first order as well. Hence m, the mean of the process, lies in $\mathcal{H}(R)$ by part (1). We thus have

$$\begin{aligned} K(s,t) &= \mathbf{E}\left(X_s X_t\right) - m(s)m(t) \\ &= \mathbf{E}\left(\langle \xi, R_s \rangle \langle \xi, R_t \rangle\right) - \langle \mathbf{E}\xi, R_s \rangle \langle \mathbf{E}\xi, R_t \rangle \\ &= \langle \Theta R_s, R_t \rangle \end{aligned}$$

for all $s, t \in T$, by Equation (2.5). Theorem 1.1 now implies that $R \geq K$ and that Θ is the corresponding dominance operator. □

The obvious converses of Theorem 2.3(1) and (2) are false, as the following examples show. Thus the way in which a random element inherits its "order" from the corresponding random process is more subtle, and will be taken up in Section 3. The relation of the Gaussianity of a random element to that of its corresponding process will discussed in Section 7.

These counterexamples also show that the necessity of the condition $R \gg K$ of our Working Conjecture (Section 1) is false.

Example. Assume the index set T to be the set of positive integers, and let the reproducing kernel R given by

$$R(s,t) = \frac{1}{st}\delta_{st}, \quad s, t = 1, 2, \dots,$$

where δ is the Kronecker delta. The corresponding RKHS $\mathcal{H}(R)$ is the set of all real sequences (a_n) such that

$$\sum_{n=1}^{\infty} n^2 a_n^2 < \infty \tag{2.7}$$

with scalar product given by

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{n=1}^{\infty} n^2 a_n b_n$$

for arbitrary elements $\mathbf{a} = (a_n)$, $\mathbf{b} = (b_n)$ of $\mathcal{H}(R)$. Addition and multiplication with a scalar in $\mathcal{H}(R)$ are defined coordinatewise. One easily verifies that $R_t \in \mathcal{H}(R)$ for all $t \in T$, and that the reproducing property holds.

An orthonormal basis in $\mathcal{H}(R)$ is given by $\mathbf{f}_n = \frac{1}{n}\mathbf{e}_n$, where

$$\mathbf{e}_n = (0, \cdots, 1, 0, \cdots), \quad n = 1, 2, \dots,$$

are the standard coordinate vectors.

We fix the sample space Ω to be the set of positive integers and the σ-algebra $\mathcal{A}$ to be the power set of Ω. Define a process X on $(\Omega, \mathcal{A})$ by

$$X_t(i) = \begin{cases} 1, & \text{if } i = t, \\ 0, & \text{if } i \neq t, \end{cases}$$

where $t \in T$ and $i \in \Omega$. We generate our counterexamples by varying the probability measure on $(\Omega, \mathcal{A})$, and in particular the mean and covariance of X. A probability measure $\mathbf{P}$ on $(\Omega, \mathcal{A})$ is of the form

$$\mathbf{P}(i) = q_i, \qquad q_i > 0, \quad \sum_{i=1}^{\infty} q_i = 1. \tag{2.8}$$

Clearly, $\{X_t, t \in T\}$ is a second-order random process with trajectories in $\mathcal{H}(R)$, having mean function $\mathbf{m}$ given by

$$\mathbf{m}(t) = \mathbf{E}[X_t] = \mathbf{P}(t) = q_t \tag{2.9}$$

and covariance function K given by

$$K(s,t) = \mathbf{E}[X_s X_t] - \mathbf{E}[X_s]\mathbf{E}[X_t] = q_t \delta_{st} - q_s q_t. \tag{2.10}$$

The random element ξ in $\mathcal{H}(R)$ defined by the process X (Equation (2.1)) is the map

$$\xi(n) = \mathbf{e}_n, \qquad n = 1, 2, \ldots .$$

The distribution $\mathbf{P}_\xi$ of ξ will be the probability measure on $(\mathcal{H}(R), \hat{\mathfrak{C}})$, where $\hat{\mathfrak{C}} = \hat{\mathfrak{C}}(\mathcal{H}(R))$ is the cylindrical σ-algebra, which is concentrated on $\{\mathbf{e}_n\}$ and such that

$$\mathbf{P}_\xi(\mathbf{e}_n) = \mathbf{P}(n) = q_n, \qquad n = 1, 2, \ldots .$$

1. Setting $q_n = C/n^{\frac{3}{2}}$, where C is the normalizing constant, we easily see that $\mathbf{m} \notin \mathcal{H}(R)$. Now Theorem 2.3 implies that ξ cannot have weak first order, although the process X has first order.

   Using Equation (2.10) to check condition (2.7), we similarly verify that $K_t \notin \mathcal{H}(R)$ for any $t \in T$. Hence $\mathcal{H}(K) \not\subseteq \mathcal{H}(R)$ (i.e., $K \not\leq R$), as claimed, even though X has its sample paths in $\mathcal{H}(R)$.
2. Let us instead put $q_n = C/n^{\frac{5}{3}}$. Then the mean $\mathbf{m}$ of X belongs to $\mathcal{H}(R)$, as is easily verified. Let ξ be the random element in $\mathcal{H}(R)$ defined by X. We will show that ξ does not have weak second order, and moreover that $K \not\leq R$.

   Consider the sequence $\mathbf{h} \in \mathcal{H}(R)$ given by $h_n = n^{-\frac{19}{12}}$, $n \in \mathbb{N}$. We have

$$\begin{aligned}\mathbf{E}\langle \xi, \mathbf{h}\rangle^2 &= \int_{\mathcal{H}(R)} |\langle \mathbf{x}, \mathbf{h}\rangle|^2 d\mathbf{P}_\xi(\mathbf{x}) \\ &= \sum_{n=1}^{\infty} \langle \mathbf{e}_n, \mathbf{h}\rangle^2 \mathbf{P}_\xi(\mathbf{e}_n) \\ &= \sum_{n=1}^{\infty} n^4 h_n^2 \frac{C}{n^{\frac{5}{3}}} = C \sum_{n=1}^{\infty} \frac{1}{n^{\frac{5}{6}}} = \infty.\end{aligned}$$

   Thus ξ does not have weak second order, although X is a second-order process. A similar argument, using the Cauchy-Schwarz inequality, shows that ξ does have weak first order.

   As in the preceding example, we see that $K \not\leq R$.

3. Driscoll's Theorem: necessity

Assume that the trajectories of a random process X belong to given RKHS $\mathcal{H}(R)$. The counterexamples above show that some additional condition is needed in order to ensure that $R \geq K$. We seek that condition in terms of the random element ξ defined by the process.

Theorem 3.1. *Let $\mathcal{H}(R)$ be a RKHS and let $\{X_t,\, t \in T\}$ be a second-order random process whose mean function and trajectories lie in $\mathcal{H}(R)$. Let K be the covariance function of X, and let ξ be the random element in $\mathcal{H}(R)$ defined by X.*

Then $R \geq K$ if and only if ξ has weak second order, in which case the covariance operator of ξ and the dominance operator L are the same.

Proof. One direction of this theorem has already been proved in Theorem 2.3.

Assume that $R \geq K$. It is easy to show that the processes X and $X - \mathbf{E}X$ have the same covariance function, and that the random elements ξ and $\xi - \mathbf{E}\xi$ have the same covariance operator. Thus, for the sake of simplicity, we will assume that X has mean zero.

Denote by L the dominance operator of $\mathcal{H}(R)$ over $\mathcal{H}(K)$. Fix an arbitrary $h \in \mathcal{H}(R)$. We want to show that $\mathbf{E}\langle \xi, h\rangle^2 < \infty$ and that L is the covariance operator of ξ. (Here all norms and inner products are in $\mathcal{H}(R)$.)

Denote by V the vector space spanned by $\{R_t,\, t \in T\}$. Let $f, g \in V$. Then there exist $t_1, \cdots, t_n \in T$ such that $f = \sum_{j=1}^n c_j R_{t_j}$, $g = \sum_{k=1}^n d_k R_{t_k}$, with some coefficients possibly zero. Then

$$\langle \xi, f\rangle = \sum c_j \langle \xi, R_{t_j}\rangle = \sum c_j X_{t_j},$$

and similarly $\langle \xi, g\rangle = \sum d_k X_{t_k}$. Therefore,

$$\begin{aligned}
\mathbf{E}\langle \xi, f\rangle \langle \xi, g\rangle &= \mathbf{E}\left[\sum_{j=1}^n c_j X_{t_j} \sum_{k=1}^n d_k X_{t_k}\right] \\
&= \sum_{j=1}^n \sum_{k=1}^n c_j d_k K(t_j, t_k) \\
&= \left\langle \sum_{j=1}^n c_j K_{t_j}, \sum_{k=1}^n d_k R_{t_k} \right\rangle \\
&= \langle Lf, g\rangle, \qquad (3.1)
\end{aligned}$$

using the reproducing property and Equation (1.6). In particular,

$$\mathbf{E}\langle \xi, f\rangle^2 = \langle Lf, f\rangle, \quad f \in V.$$

Fix a sequence (h_n) in V such that $\lim_{n\to\infty} \|h - h_n\| = 0$ (Lemma 4.1), and consider the random variables $Y_n = \langle \xi, h_n\rangle$. By the above, each Y_n is a second-order random variable with $\mathbf{E}Y_n^2 = \langle Lh_n, h_n\rangle$ and

$$\mathbf{E}\,[Y_m Y_n] = \langle Lh_m, h_n\rangle.$$

Consequently, taking into an account continuity of the dominance operator L, we have

$$\lim_{m\to\infty} \lim_{n\to\infty} \mathbf{E}\,[Y_m Y_n] = \langle Lh, h\rangle.$$

independent of the way $m, n \to \infty$. This, in turn, implies the mean-square convergence of the sequence (Y_n) [12, Corollary to Theorem 2.4.5]. That is, there exists a second-order random variable Y such that $\mathbf{E}Y^2 = \langle Lh, h\rangle$ and $Y_n \to Y$ in the mean square sense. Thus, there exists a subsequence, which we again call Y_n, such that $Y_n \to Y$ a.s.

On the other hand, $h_n \to h$ in $\mathcal{H}(R)$ implies

$$Y_n = \langle \xi, h_n\rangle \to \langle \xi, h\rangle$$

everywhere on Ω. Therefore, $Y = \langle \xi, h \rangle$ a.s. In other words,

$$\mathbf{E}\langle \xi, h \rangle^2 = \langle Lh, h \rangle \tag{3.2}$$

for all $h \in \mathcal{H}(R)$. This shows that ξ has a weak second order, and thus has a covariance operator.

Using Equation (3.2) and the parallelogram law, one easily extends Equation (3.1) to all $f, g \in \mathcal{H}(R)$. Hence, by Equation (2.5), L is the covariance operator of ξ. $\square$

Replacing dominance by nuclear dominance in the conclusion of Theorem 3.1 introduces some further subtleties. Suppose that $\xi : \Omega \to H$ and $\eta : \Omega \to H$ are two random elements in a Hilbert space H. We say that η is a *version* of ξ if

$$\mathbf{P}\left(\langle \xi, h \rangle = \langle \eta, h \rangle\right) = 1$$

for all $h \in H$. It is easy to see that if η is a version of ξ then they define the same mean and the same covariance operator, provided these exist. Separably valued random elements were introduced in Definition 2.1.

Theorem 3.2. *Let H be a Hilbert space, and let μ be a probability measure on the cylindrical σ-algebra $\hat{\mathfrak{C}}(H)$. Then μ has a unique Radon extension to the Borel σ-algebra $\mathfrak{B}(H)$.*

Equivalently, every random element ξ in H has a separably valued version.

The probability distribution of a separably valued Borel random element in a Hilbert space is a Radon probability measure.

Proof. Since every Hilbert space is a reflexive Banach space, the existence of a Radon extension follows from [16, Corollary 5 to Theorem I.5.3].

The existence of a Radon extension is equivalent to the existence of a separably valued version of ξ [16, Theorem IV.2.7].

If ξ is separably valued, then ξ is a Borel random element [16, Theorem II.1.1], so that the closure Y of $\xi(\Omega)$ is both Borel and a separable subset of H. Thus, $\mathbf{P}_\xi(Y) = \mathbf{P}(\Omega) = 1$, and so $\mathbf{P}_\xi$ is Radon [16, Corollary to Theorem I.3.1]. $\square$

We may now restate Theorem 3.1 in the case of nuclear dominance.

Corollary 3.1. *Let $\mathcal{H}(R)$ be a RKHS, and let $\{X_t, t \in T\}$, K and ξ be as in Theorem 3.1.*

Then $R \gg K$ if and only if ξ has a separably valued version η with strong second order, in which case the covariance operator of ξ and the dominance operator L are the same.

Proof. It follows from Theorem 3.2 that ξ always has a separably valued version η. We first assume that η has strong second order and prove that $R \gg K$.

The fact that $R \geq K$ follows from Theorem 3.1 applied to η. The probability distribution $\mathbf{P}_\eta$ is a Radon measure because η is separably valued (Theorem 3.2). But covariance operators of strong second order Radon probability measures in Hilbert space are nuclear [16, Corollary to Proposition III.2.3]. Hence $R \gg K$.

Conversely, suppose that $R \gg K$, with (nuclear) dominance operator L. Since $R \geq K$, Theorem 3.1 implies that L is the covariance operator of ξ. Let η be a separably valued version of ξ as established by Theorem 3.2. Since L is also the covariance operator of η, and since L is nuclear, it follows, again by [16, Corollary to Proposition III.2.3], that η has strong second order. $\square$

Note that the introduction of the version η would not be necessary if $\mathcal{H}(R)$ were separable.

Remark 3.1. Corollary 3.1 shows that $R \gg K$ is a necessary condition for sample paths of X to belong to $\mathcal{H}(R)$, *provided* that the corresponding random element ξ has strong second order. While Fortet did not include such a hypothesis in [6, Theorem 2], he did use the strong second order assumption in the actual proof, as revealed in his unpublished notes [7]. The present results were arrived at independently.

Having examined the necessity of nuclear dominance, our next goal is to study its sufficiency. We first must pause to develop the mathematical machinery we will need.

4. Some results on RKHS

In this section we will develop most of the properties of reproducing kernel Hilbert spaces that we will need, culminating in an extension of Driscoll's trace formula (1.1) (Proposition 4.5). The properties of the RKHS $\mathcal{H}(R,T)$ depend on the kernel R and the set T, and in this section we will be allowing both to vary. As mentioned in the Introduction, we will sometimes index the norm and inner product of $\mathcal{H}(R,T)$ by R, sometimes by T, and once in a while by both. When T is understood, we will write $\mathcal{H}(R)$ for $\mathcal{H}(R,T)$. Hopefully this will maximize clarity while minimizing notation.

Lemma 4.1. *Let $R : T \times T \mapsto \mathbb{R}$ be a reproducing kernel. The set $\{R_t\,, t \in T\}$ is total in $\mathcal{H}(R)$, and the vector space V spanned by $\{R_t, t \in T\}$ is dense in $\mathcal{H}(R)$.*

See [17, Theorem 5B].

The function

$$d_R(s,t) = \|R_s - R_t\|$$

defines a pseudo-metric on T. Note that $d_R^2(s,t) = R(s,s) - 2R(s,t) + R(t,t)$. Observing that $\sum\sum a_i a_j R(t_i,t_j) = \|\sum a_i R_{t_i}\|^2$, we have the following.

Lemma 4.2. *R is nonsingular on T iff the set $\{R_t, t \in T\}$ is linearly independent. If R is nonsingular on T, then d_R is a metric on T.*

Note that d_R may still be metric even if the covariance kernel R is singular. For example, let $T = \{1,2,3\}$ and let $R(s,t)$ be given by the elements of the matrix

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix}.$$

It is easy to check that $d_R(1,2) = 1, d_R(1,3) = \sqrt{2}$, and $d_R(2,3) = 1$. Thus the converse of the second statement in Lemma 4.2 is false.

Lemma 4.3. *Let T be an arbitrary set and R a covariance kernel on T. Assume that d_R is a metric on T. Then*

1. *every $f \in \mathcal{H}(R)$ is d_R-continuous, and*
2. *the metric space (T, d_R) is separable iff $\mathcal{H}(R)$ is separable.*

Proof. The first conclusion follows from the fact that $|f(s)-f(t)| = |\langle f, R_s - R_t\rangle| \leq \|f\| d_R(s,t)$, using the reproducing property and the Cauchy-Schwarz inequality.

Assume (T, d_R) is a separable metric space, and let $S \subset T$ be countable and d_R-dense in T. Denote by V the linear span of $\{R_t, t \in T\}$, and by W the linear span of $\{R_t, t \in S\}$ over the field of rationals. V is dense in $\mathcal{H}(R)$ by Lemma 4.1. Clearly, W is dense in V, hence in $\mathcal{H}(R)$, and is countable.

Now assume that $\mathcal{H}(R)$ is separable. Then for any $\epsilon > 0$ there exists a countable partition $\{T_j(\epsilon)\}$ of T such that each set $T_j(\epsilon)$ has d_R-diameter smaller than ϵ [5, Theorem 1.2]. Picking a point in each $T_j(\epsilon)$ for each $\epsilon = 1, \frac{1}{2}, \frac{1}{3}, \ldots$, one obtains a countable, d_R-dense subset of T. ☐

Remark 4.1. Borrowing an idea of Dudley, Talagrand [21, pp. 101, 116] defined a *canonical metric* on T by

$$d(s,t) = (\mathbf{E}(X_s - X_t)^2)^{1/2}, \qquad s, t \in T,$$

where $\{X_t, t \in T\}$ is a second-order, zero-mean process. If K is the covariance of the process, then the Loève isometry gives

$$d(s,t) = \|K_s - K_t\| = d_K(s,t).$$

Independently, Fortet [5] introduced the function $\Delta(s,t) = R(s,s) - 2R(s,t) + R(t,t)$ just before stating his Theorem 1.2, which we have quoted in the preceding proof. Of course, $\Delta(s,t) = d_R^2(s,t)$. This definition makes no reference to any particular stochastic process, which is what makes it useful in our application. For us the kernel R that defines d_R need not be the same as the covariance K of the process we are given.

In order to deal with singular kernels, we introduce next the notion of an *R-Hamel subset* of T. Recall that a set $\{v_\alpha\}$ of linearly independent vectors in a vector space V is called a *Hamel basis* (in V) if $\{v_\alpha\}$ spans V. By the Hausdorff Maximality Principle, every vector space has a Hamel basis. Each vector $v \in V$ has a unique representation as a finite linear combination of vectors from the Hamel basis $\{v_\alpha\}$. It is easily shown that if V is a linear space and $W \subset V$ spans V, then there exists a Hamel basis B of V such that $B \subseteq W$.

Definition 4.1. Let V be the vector space spanned by $\{R_t, t \in T\}$. A set $T_0 \subseteq T$ such that $\{R_t,\, t \in T_0\}$ is a Hamel basis of V will be called an *R-Hamel subset of* T, or just a *Hamel subset* of T when it is clear which reproducing kernel is in question.

If $\mathcal{H}(R)$ has infinite (orthogonal) dimension, then one can show that a Hamel subset of T must be infinite (and in fact uncountable).

Example. Consider a standard Wiener process $\{W_t,\, t \in [0,1]\}$ and its covariance $K(t,s) = \min(t,s)$. Let n and $0 < t_1 < \cdots < t_n \leq 1$ be arbitrary. Denote by K_n the matrix obtained by restricting the covariance K to the set $\{t_1, \cdots, t_n\}$. It is not difficult to show that

$$\det K_n = \prod_{k=1}^{n} (t_k - t_{k-1})\,; \quad t_0 \equiv 0.$$

Thus K_n is invertible, so $K_{t_1}, \cdots, K_{t_n}$ are linearly independent. Since $K_0 \equiv 0$, we have $T_0 = (0,1]$ in this case.

For arbitrary kernels, we have

Lemma 4.4. *If T_0 is a Hamel subset of T, then $\{R_t, t \in T_0\}$ is total in $\mathcal{H}(R,T)$, and (T_0, d_R) is a metric space.*

Fortet [5, Theorem 1.1] gives a useful criterion for membership in a RKHS $\mathcal{H}(R,T)$. Let $\mathcal{F} = \{S \subset T : S \text{ is finite}\}$.

Theorem 4.1. *Let $f : T \to \mathbb{R}$ be a function. Then $f \in \mathcal{H}(R,T)$ iff*

$$\sup_S \sup_{a_i} \frac{|\sum_i a_i f(t_i)|^2}{\sum_i \sum_j a_i a_j R(t_i,t_j)} < \infty, \tag{4.1}$$

where the suprema are taken over all $S = \{t_1, \dots, t_n\} \in \mathcal{F}$ and all real $a_1, \dots, a_n$, with n arbitrary, such that the denominator in (4.1) *is not zero. When the inequality* (4.1) *holds, the left-hand-side is $\|f\|^2$ in $\mathcal{H}(R,T)$.*

If the set T itself is finite, then $\|f\|^2$ is just the inner supremum of (4.1). An application of Cauchy-Schwarz and the fact that a symmetric matrix has a square root yield the following well-known result:

Lemma 4.5. *Let T be finite, and let the matrix determined by a reproducing kernel R be nonsingular. Then, for all $f, g \in \mathcal{H}(R)$:*

$$\|f\|_R^2 = \sum_{t,s \in T} f(t)f(s)R^{-1}(t,s),$$

$$\langle f, g\rangle_R = \sum_{t,s \in T} f(t)g(s)R^{-1}(t,s),$$

where R^{-1} is the inverse of the matrix R.

This is a standard result in the theory of RKHS. See, for example, [1, p. 346].

The restriction of a covariance kernel to a subset S of T is still a covariance kernel, and defines a RKHS $\mathcal{H}(R,S)$. Despite the abuse of notation, we will use $\|\cdot\|_T$, etc., when it is important to indicate the underlying index set. We let f_S denote the restriction of a function to S. For simplicity, we will write $\|f\|_S$ for $\|f_S\|_S$. This is justified by the following result.

Proposition 4.1. *Let $\mathcal{H}(R,T)$ be a RKHS with kernel R, and $S \subseteq T$. Then the restriction of R to S is a covariance kernel. If $f \in \mathcal{H}(R,T)$, then $f_S \in \mathcal{H}(R,S)$ and*

$$\|f\|_S \le \|f\|_T. \tag{4.2}$$

The map $J : \mathcal{H}(R,T) \to \mathcal{H}(R,S)$ given by $J(f) = f_S$ has the following properties:

1. *J maps $\mathcal{H}(R,T)$ linearly onto $\mathcal{H}(R,S)$.*
2. *The nullspace of J is the subspace*

$$F_0 = \{h \in \mathcal{H}(R,T) : h \text{ vanishes on } S\}. \tag{4.3}$$

3. *Let F_1 be the orthocomplement of F_0 in $\mathcal{H}(R,T)$. Then F_1 is generated by the family $\{R_t, t \in S\}$, and J maps F_1 isometrically onto $\mathcal{H}(R,S)$.*

The properties of f_S, including inequality (4.2), follow easily from criterion (4.1). The properties of the map J may be found in [15, Proposition 3.15] or in [1, pp. 351ff].

We say that $S \subseteq T$ is a *determining set for* $\mathcal{H}(R,T)$ if the only function $f \in \mathcal{H}(R,T)$ which vanishes on S is the zero function. For example, a Hamel subset of T is a determining set (Lemma 4.4).

From Proposition 4.1 we have the following.

Proposition 4.2. *Let $S \subseteq T$, and let J be the restriction map $f \mapsto f_S$. The following are equivalent:*

1. *S is a determining set for $\mathcal{H}(R,T)$.*
2. *J in an isometry between $\mathcal{H}(R,T)$ and $\mathcal{H}(R,S)$.*
3. *Every $f \in \mathcal{H}(R,S)$ has a unique extension to an element $f \in \mathcal{H}(R,T)$, and $\|f\|_T = \|f\|_S$.*
4. *The family $\{R_t, t \in S\}$ is total in $\mathcal{H}(R,T)$.*
5. *The family of linear functionals on $\mathcal{H}(R,T)$ determined by $\{R_t, t \in S\}$ separates points of $\mathcal{H}(R,T)$.*

The concept of a determining set is due to Fortet [6]. His definition is actually condition (2).

Proposition 4.3. *Suppose $\mathcal{H}(R,T)$ is separable, and let T_0 be a Hamel subset of T. Then (T_0, d_R) is a separable metric space. If S_0 is a countable d_R-dense subset of T_0, then S_0 is a determining set for $\mathcal{H}(R,T)$.*

Proof. Separability of $\mathcal{H}(R,T)$ and Proposition 4.1 imply that $\mathcal{H}(R,T_0)$ is separable. Lemma 4.3 now implies that (T_0, d_R) is a separable metric space.

Let S_0 be a countable d_R-dense subset of T_0, and let f be an element of $\mathcal{H}(R,T)$ which vanishes on S_0. As an element of $\mathcal{H}(R,T_0)$, the restriction f_{T_0} is d_R-continuous and vanishes on S_0, so must vanish on T_0 as well. But T_0 is a determining set for $\mathcal{H}(R,T)$, so f must vanish on T. □

In addition to the notational conventions adopted earlier, we may write $\|f\|_S$ for an arbitrary function $f : T \to \mathbb{R}$, with the understanding that this means $\|f_S\|_S$ and presumes that $f_S \in \mathcal{H}(R,S)$.

Lemma 4.6. *Assume $\mathcal{H}(R,T)$ separable, and let $S_0 = \{s_1, s_2, \cdots\}$ be a countable d_R-dense subset of a Hamel subset T_0 in T. Put $S_n = \{s_1, \cdots, s_n\}$. Then, for each $f \in \mathcal{H}(R,T)$, $\|f\|_{S_n}$ is monotone increasing, and*

$$\|f\|_T = \lim_{n\to\infty} \|f\|_{S_n}. \tag{4.4}$$

Conversely, if a function $f : T \mapsto \mathbb{R}$ satisfies

$$\lim_{n\to\infty} \|f\|_{S_n} < \infty, \tag{4.5}$$

then $f_{S_0} \in \mathcal{H}(R,S_0)$ and there exists a unique $f_1 \in \mathcal{H}(R,T)$ that coincides with f on S_0, and

$$\|f_1\|_T = \|f\|_{S_0} = \lim_{n\to\infty} \|f\|_{S_n}. \tag{4.6}$$

Proof. Let $f \in \mathcal{H}(R,T)$. The fact that the sequence $\|f\|_{S_n}$ is increasing follows from (4.2). By Proposition 4.3 it suffices to prove (4.4) when $T = S_0$, that is, when T is countable. But this follows easily from condition (4.1), as we note that any finite subset S of T is eventually contained in some set S_n.

Conversely, if (4.5) holds, then a similar argument shows that the restriction f_{S_0} belongs to $\mathcal{H}(R,S_0)$, and by Proposition 4.3 it has a unique extension to a function $f_1 \in \mathcal{H}(R,T)$. Now (4.6) follows from Proposition 4.2(3) and (4.4). □

Remark 4.2. If in addition d_R is a metric on T and f is d_R-continuous, then it is easy to see that f coincides with f_1 on T_0. However, unless R is nonsingular, this is not enough to conclude that $f \in \mathcal{H}(R,T)$, since we do not know whether S_0 is necessarily dense in T.

We have discussed the relation between $\mathcal{H}(R,T)$ and the RKHS derived from it by changing the kernel R to a "smaller" kernel K, or by changing the set T to a subset S. We now begin to combine these ideas. Define F_0 as in (4.3), with orthocomplement (in $\mathcal{H}(R,T)$) denoted F_1, and similarly define $H_0 = \{h \in \mathcal{H}(K,T) : h \text{ vanishes on } S\}$, with orthocomplement (in $\mathcal{H}(K,T)$) denoted H_1. Assume $\mathcal{H}(K,T) \subseteq \mathcal{H}(R,T)$. Clearly $H_0 = F_0 \cap \mathcal{H}(K,T)$.

Lemma 4.7. *Let $\mathcal{H}(K,T) \subseteq \mathcal{H}(R,T)$, with dominance map L, and let $S \subseteq T$. Then*

1. *$\mathcal{H}(K,S) \subseteq \mathcal{H}(R,S)$.*
2. *L maps F_1 into H_1.*

Proof. (1) Let $\hat{f} \in \mathcal{H}(K,S)$. By Proposition 4.1(1), $\hat{f} = f_S$ for some $f \in \mathcal{H}(K,T)$. But then $f \in \mathcal{H}(R,T)$, so $\hat{f} \in \mathcal{H}(R,S)$.

(2) Let $f \in F_1$. We claim that $Lf \perp H_0$ (with respect to the inner product in $\mathcal{H}(K,T)$). To this end, let $g \in H_0$. Then $g \in F_0$, so (using (1.5)) $\langle Lf, g\rangle_K = \langle f, g\rangle_R = 0$. □

Our next goal is to extend Driscoll's *trace formula* (1.1) to a much more general setting. Namely, we consider an arbitrary set T, an arbitrary reproducing kernel R on T, and an arbitrary (nuclear) dominance operator L on $\mathcal{H}(R,T)$. Part 1 of Lemma 4.7 leads us to introduce a second dominance map, whose relation to L is given in the next result. The announced generalization of (1.1) is then achieved by Proposition 4.5.

Proposition 4.4. *Suppose that $\mathcal{H}(K,T) \subseteq \mathcal{H}(R,T)$, with dominance map L, and $S \subseteq T$. Let F_1 and H_1 be as defined above, and let L_S be the dominance map of $\mathcal{H}(R,S)$ over $\mathcal{H}(K,S)$. Finally, let J^{-1} be the inverse of the restriction of J to F_1. Then*

$$L_S = JLJ^{-1}. \tag{4.7}$$

If L is nuclear then so is L_S, and

$$\mathrm{Tr}(L_S) \le \mathrm{Tr}(L), \tag{4.8}$$

with equality if S is a determining set for $\mathcal{H}(R,T)$.

The proposition may be visualized with the following diagram:

$$\begin{tikzcd} F_1 \arrow[r, "L"] \arrow[d, "J"'] & H_1 \arrow[d, "J"'] \\ \mathcal{H}(R,S) \arrow[r, "L_S"] & \mathcal{H}(K,S) \end{tikzcd}$$

Here L_S is the dominance map of $\mathcal{H}(R,S)$ over $\mathcal{H}(K,S)$, as guaranteed by Lemma 4.7. Invertibility of J when restricted to F_1 follows from Proposition 4.1. Equation (4.7) says in effect that the diagram commutes.

Equation (4.7) appears in Fortet's unpublished notes [7] as part of his proof of inequality (4.8), which is announced without proof in [6, Theorem 3(1)]. The equation was discovered independently as part of the present research, and the proof below is a bit different from Fortet's.

Proof. In the following, we will index inner products by both a kernel (R or K) and a set (T or S).

Let $\hat{L} = JLJ^{-1}$. It is clear that $\hat{L}$ is a linear map from $\mathcal{H}(R,S)$ to $\mathcal{H}(K,S)$. Thus, according to Theorem 1.1, to show that $\hat{L}$ is the dominance map L_S we must show that $\langle \hat{L}\hat{f}, \hat{g}\rangle_{K,S} = \langle \hat{f}, \hat{g}\rangle_{R,S}$ for all $\hat{f} \in \mathcal{H}(R,S)$ and $\hat{g} \in \mathcal{H}(K,S)$.

Fixing $\hat{f}$ and $\hat{g}$, there are unique functions $f \in F_1$ and $g \in H_1$ such that $\hat{f} = Jf$ and $\hat{g} = Jg$. Now $Lf \in H_1$ by Lemma 4.7, and since J acts as an isometry on H_1 and F_1, and L is a dominance map, we have

$$\begin{aligned} \langle \hat{L}\hat{f}, \hat{g}\rangle_{K,S} &= \langle JLf, Jg\rangle_{K,S} = \langle Lf, g\rangle_{K,T} \\ &= \langle f, g\rangle_{R,T} = \langle Jf, Jg\rangle_{R,S} = \langle \hat{f}, \hat{g}\rangle_{R,S}, \end{aligned}$$

as desired. This establishes Equation (4.7).

From (4.7) and the isometric property of J we readily see that we have

$$\langle L_S Jf, Jg\rangle_{R,S} = \langle Lf, g\rangle_{R,T} \quad \text{for all } f, g \in F_1. \tag{4.9}$$

Now let $\{\hat{e}_k\}$ (not necessarily countable) be a CON (complete orthonormal) set in $\mathcal{H}(R,S)$. Then the functions $e_k = J^{-1}\hat{e}_k$ form a CON set in F_1, and from (4.9) we have $\langle L_S\hat{e}_k, \hat{e}_k\rangle_{R,S} = \langle Le_k, e_k\rangle_{R,T}$ for each k. If L is nuclear, then summing over k gives (4.8). Moreover, when S is a determining set for $\mathcal{H}(R,T)$ then $\mathcal{H}(R,T) = F_1$ and so we have equality in (4.8). $\square$

Proposition 4.5. *Let $\mathcal{H}(R,T)$ be separable and let the kernel K be such that $R \gg K$, with dominance operator L. Let T_0 be an R-Hamel subset of T and $S_0 = \{s_1, s_2, \cdots\}$ be a d_R-dense subset of T_0. Denote by K_n and R_n the matrices obtained by restricting the kernels K and R to the set $\{s_1, \cdots, s_n\} \subset S_0$. Then*

$$\operatorname{Tr}(L) = \lim_{n\to\infty} \operatorname{Tr}\left(K_n R_n^{-1}\right). \tag{4.10}$$

Proof. S_0 is a determining set for $\mathcal{H}(R,T)$ by Proposition 4.3, so Proposition 4.4 implies that $\operatorname{Tr}(L) = \operatorname{Tr}(L_{S_0})$, where L_{S_0} is the dominance operator for $\mathcal{H}(R,S_0)$ over $\mathcal{H}(K,S_0)$. Thus we are reduced to proving the proposition when T is a countable set, which we continue to denote by $\{s_1, s_2, \dots\}$.

First we apply the Gram-Schmidt procedure to the functions $R_t, t \in T$, to get a CON basis $e_1, e_2, \dots$ of $\mathcal{H}(R,T)$. Of course, $\operatorname{Tr}(L) = \sum_i \langle Le_i, e_i\rangle_{R,T}$.

Next, let $S_n = \{s_1, \dots, s_n\}$, let L_n be the dominance operator of $\mathcal{H}(R,S_n)$ over $\mathcal{H}(K,S_n)$, and let $J : f \mapsto f_{S_n}$ be the restriction map (J also depends on n). Defining the subspace $F_1 \subseteq \mathcal{H}(R,T)$ as in Proposition 4.1, we see that F_1 is generated by the functions $R_{s_1}, \dots R_{s_n}$, and therefore by $e_1, \dots, e_n$, and that the images $\hat{e}_i = Je_i$ are an orthonormal basis of $\mathcal{H}(R,S_n)$. Thus $\operatorname{Tr}(L_n) = \sum_i^n \langle L_n\hat{e}_i, \hat{e}_i\rangle_{R,S_n} = \sum_i^n \langle Le_i, e_i\rangle_{R,T}$, so

$$\lim_n \operatorname{Tr}(L_n) = \operatorname{Tr}(L). \tag{4.11}$$

Finally, if we view L_n as a linear map between finite-dimensional vector spaces, then property (1.6) means that for each $t \in S_n$, L_n maps $(R(s_1,t), \dots, R(s_n,t))$ to $(K(s_1,t), \dots, K(s_n,t))$. These vectors are columns of the matrices R_n and K_n, respectively. Therefore L_n, viewed as a matrix, satisfies $L_nR_n = K_n$, and since R is nonsingular, $L_n = K_nR_n^{-1}$. From this and (4.11) we have Equation (4.10). $\square$

In this proof we have adapted some ideas of Driscoll [3, proof of Theorem 3]. An alternate, perhaps more elegant approach rests on Fortet's refinement of inequality (4.8) in [6, Theorem 3(2)], which asserts that

$$\operatorname{Tr}(L) = \sup_S \operatorname{Tr}(L_S), \tag{4.12}$$

the supremum taken over all finite $S \subset T$. After reducing to the case that T is countable, we may argue as in Lemma 4.6 to replace the right-hand side of (4.12) by the supremum over the sets S_n, leading immediately to Equation (4.11). We note the similarity of Equation (4.12) to Fortet's criterion (4.1).

5. Driscoll's Theorem: sufficiency

We are ready to address a key problem: Given a second order process X with covariance K, and a RKHS $\mathcal{H}(R)$, when can we say that

$$\mathbf{P}\left(X_{\boldsymbol{\cdot}} \in \mathcal{H}(R)\right) = 1? \tag{5.1}$$

Driscoll proved (5.1) assuming that nuclear dominance ($R \gg K$) holds and that X is Gaussian, along with certain other conditions (see Section 1). We will show below that nuclear dominance alone is sufficient for the RKHS $\mathcal{H}(R)$ to contain the sample paths of a version of X almost surely. As our counterexamples in Section 2 show, $R \gg K$ is not a necessary condition for (5.1), nor is there a hope of achieving a similar generalization for the necessary condition in terms of ordinary dominance only.

The present proof is inspired by Driscoll's [3, p. 313], but is necessarily longer due to our abandoning any assumptions on the set T or the kernels R and K. (Driscoll himself does not use the Gaussian assumption in this part of his proof.)

Theorem 5.1. *Let $X = \{X_t,\, t \in T\}$ be a real second-order stochastic process with covariance K, and let R be another covariance kernel on T such that $R \gg K$. Assume that the mean function of the process belongs to $\mathcal{H}(R)$, as well.*

Then there exists a version Y of X whose trajectories belong to $\mathcal{H}(R)$ almost surely.

If in addition d_R is a metric, $\mathcal{H}(R)$ is separable, and the trajectories of X are d_R-continuous, then the trajectories of X belong to $\mathcal{H}(R)$ almost surely with respect to the completion measure.

Proof. It is not hard to show that $R \gg K$ implies the existence of a kernel R_1 such that $R \geq R_1 \gg K$ and such that $\mathcal{H}(R_1)$ is separable [6, Theorem 1]. Therefore, without loss of generality, we will just assume that $\mathcal{H}(R)$ is separable. Similarly, upon considering the process $X - \mathbf{E}X$, we may assume $\mathbf{E}X = 0$. Finally, we will denote by $(\Omega, \mathcal{A}, \mathbf{P})$ the probability space on which the process is defined.

Fix an R-Hamel subset T_0, and a countable set $S_0 = \{s_1, s_2, \dots\}$ which is d_R-dense in T_0. The existence of S_0 follows from Proposition 4.3. We will prove first that the process $\{X_s,\, s \in S_0\}$ has trajectories in the RKHS $\mathcal{H}(R, S_0)$ with probability one.

Indeed, for each $n = 1, 2, \dots$, let

$$S_n = \{s_1, s_2, \dots, s_n\}.$$

Denote the restrictions of K, R on S_n by K_n, R_n, respectively, and using Lemma 4.5, define

$$Z_n = \|X\|^2_{S_n} = \sum_{i=1}^{n}\sum_{j=1}^{n} X_{s_i}X_{s_j}R_n^{-1}\left(s_i, s_j\right), \quad n = 1, 2, \ldots.$$

Clearly, Z_n is a random variable. The sequence $\{Z_n\}$ is non-decreasing, by Lemma 4.6, and so defines a random variable

$$Z = \lim_{n\to\infty} Z_n$$

with the possibility that Z can take infinite values. But now, as Driscoll [3] argues, the Monotone Convergence Theorem implies that

$$\begin{aligned}\mathbf{E}[Z] &= \lim_{n\to\infty}\mathbf{E}[Z_n] \\ &= \lim_{n\to\infty}\sum_{i=1}^{n}\sum_{j=1}^{n} K_n\left(s_i, s_j\right)R_n^{-1}\left(s_i, s_j\right) \\ &= \lim_{n\to\infty}\operatorname{Tr}(K_nR_n^{-1}) \\ &= \operatorname{Tr}(L) < \infty,\end{aligned}$$

by Proposition 4.5. Hence $\mathbf{P}(Z < \infty) = 1$. In other words, there exists a set $\Omega' \in \mathcal{A}$ with $\mathbf{P}(\Omega') = 1$ and such that $Z(\omega) < \infty$ for all $\omega \in \Omega'$. Thus, for each $\omega \in \Omega'$, the corresponding trajectory $X_{\bullet}(\omega)$ of the process $\{X_s,\, s \in S_0\}$ belongs to the RKHS $\mathcal{H}(R, S_0)$, and defines a unique function $t \mapsto f(\omega, t) \in \mathcal{H}(R)$, by Lemma 4.6. Define

$$\xi(\omega) = \begin{cases} f(\omega, \cdot), & \omega \in \Omega', \\ 0, & \text{otherwise.}\end{cases}$$

Claim. *ξ is a random element in $\mathcal{H}(R)$ and has strong second order and mean zero.*

Proof of the claim. First we need to prove that ξ is a Borel random element. Indeed, $\mathcal{H}(R)$ is assumed separable. For every $s \in S_0$, $\langle \xi, R_s\rangle = X_s$ is a random variable. Moreover, the family $\{R_s,\, s \in S_0\}$ separates points of $\mathcal{H}(R)$ by Proposition 4.2. Thus ξ is Borel by [16, Theorem II.1.1]. Consequently, $\|\xi\|$ is measurable, and since

$$\|\xi\|^2_{S_n} = \|X\|^2_{S_n} = Z_n,$$

Lemma 4.6 and Proposition 4.5 show that

$$\begin{aligned}\mathbf{E}\|\xi\|^2 &= \mathbf{E}\left[\lim_{n\to\infty}\|\xi\|^2_{S_n}\right] \\ &= \mathbf{E}\left[\lim_{n\to\infty} Z_n\right] \\ &= \lim_{n\to\infty}\mathbf{E}\left[Z_n\right] \\ &= \operatorname{Tr}(L) \;<\; \infty,\end{aligned}$$

which proves that ξ has strong second order.

Of course, strong second order implies weak first order. Hence, by Theorem 2.1, ξ has expectation $\mathbf{E}\xi \in \mathcal{H}(R)$. Using Equation (2.3), we have

$$\langle \mathbf{E}\xi, R_s\rangle = \mathbf{E}\langle \xi, R_s\rangle = 0$$

for all $s \in S_0$. Hence, $\mathbf{E}\xi = 0$ by Proposition 4.3. This proves the claim.

Now let $\{Y_t,\, t \in T\}$ be the process defined by the random element ξ. Then Y is a zero-mean, second-order process by Theorem 2.3, with trajectories in $\mathcal{H}(R)$.

We next show that $\mathbf{P}(Y_t = X_t) = 1$, for all $t \in T$. To do so, we will first show that

$$\|Y_t - X_t\|_{L^2}^2 \equiv \mathbf{E}\left(Y_t - X_t\right)^2 = 0, \quad \forall t \in T_0. \tag{5.2}$$

Note first that (5.2) is true for $t \in S_0$, by definition of Y.

Fix arbitrary $t \in T_0 \setminus S_0$. Then, for any $s \in S_0$,

$$\|X_t - Y_t\|_{L^2} \le \|X_t - X_s\|_{L^2} + \|X_s - Y_s\|_{L^2} + \|Y_s - Y_t\|_{L^2}$$

The middle term is zero. As to the first term, Loève's isometry and Equation (1.5) yield

$$\begin{aligned}
\|X_t - X_s\|_{L^2}^2 &= \|K_t - K_s\|_K^2 \\
&= \langle K_t - K_s, K_t - K_s\rangle_K \\
&= \langle L(R_t - R_s), L(R_t - R_s)\rangle_K \\
&= \langle L(R_t - R_s), R_t - R_s\rangle_R \\
&\le \|L\| d_R^2(t,s).
\end{aligned}$$

The third term above is handled by noting that

$$\begin{aligned}
\|Y_s - Y_t\|_{L^2}^2 &= \mathbf{E}\left(Y_s - Y_t\right)^2 \\
&= \mathbf{E}\langle \xi, R_s - R_t\rangle_R^2 \\
&\le \mathbf{E}\|\xi\|_R^2 d_R^2(s,t) = \operatorname{Tr}(L) d_R^2(s,t).
\end{aligned}$$

Thus, for $t \in T_0 \setminus S_0$ and $s \in S_0$,

$$\|X_t - Y_t\|_{L^2} \le \left(\sqrt{\|L\|} + \sqrt{\operatorname{Tr}(L)}\right) d_R(t,s),$$

which can be made arbitrary small, and (5.2) holds. Since $\mathbf{E}X_t = \mathbf{E}Y_t$ for all t, it follows that $Y_t = X_t$ a.s. for all $t \in T_0$.

Finally, take $t \in T \setminus T_0$. Then, in a unique way, $R_t = \sum_{k=1}^n a_k R_{t_k}$ for some $t_k \in T_0$, $a_k \in \mathbb{R}$. Thus $K_t = LR_t = \sum_1^n a_k K_{t_k}$, so by Loève's isometry $X_t = \sum_1^n a_k X_{t_k}$ a.s. On the other hand,

$$Y_t = \langle \xi, R_t\rangle_R = \sum_1^n a_k \langle \xi, R_{t_k}\rangle_R = \sum_1^n a_k Y_{t_k}.$$

Since $X_{t_k} = Y_{t_k}$ a.s., we have $X_t = Y_t$ a.s.

It remains to prove that under the additional assumptions of separability of $\mathcal{H}(R)$ and d_R-continuity of the sample paths of X, those sample paths belong to $\mathcal{H}(R)$ almost surely, this time with respect to the completion of $\mathbf{P}$.

By Lemma 4.3, the trajectories of Y are d_R-continuous and (T, d_R) is separable. Denote by $T_d \subset T$ a countable d_R-dense set. Since Y is a version of X, there exists a set $\Omega_1 \subset \Omega$, with $\mathbf{P}(\Omega_1) = 1$, and such that

$$X_t(\omega) = Y_t(\omega), \quad \text{for all } \omega \in \Omega_1 \text{ and all } t \in T_d.$$

Both processes X and Y have d_R-continuous sample paths that coincide, at least for $\omega \in \Omega_1$, on a countable d_R-dense set T_d; hence, they coincide on T. Therefore, denoting the completion of $\mathbf{P}$ still by $\mathbf{P}$,

$$\mathbf{P}(X_{\bullet} \in \mathcal{H}(R)) \geq \mathbf{P}(X \equiv Y) \geq \mathbf{P}(\Omega_1) = 1.$$

The theorem is proved. □

Remark 5.1. The final part of the proof of Theorem 5.1 required the use of a countable dense set T_d in T. We already had established that $X_t = Y_t$ for all t in the countable set S_0. But we only know that S_0 is dense in the Hamel subset T_0 of T, and as pointed out in Remark 4.2 this approach would only allow us to conclude that $X_t = Y_t$ for all t in T_0.

6. Existence: sufficiency

We have just seen that the condition $R \gg K$ is sufficient for a *given* process X of covariance K (or at least a version of X) to have its sample paths in $\mathcal{H}(R)$. As we noted in the Introduction, Fortet asserted (correctly) that this condition is sufficient for the *existence* of such a process X. On the other hand, we now know that this condition is not necessary, and so we may ask whether we may substitute a weaker sufficient condition for existence of such a process.

Theorem 6.1. *Let $\mathcal{H}(R)$ and $\mathcal{H}(K)$ be two RKHS such that $\mathcal{H}(K)$ is a separable subset of $\mathcal{H}(R)$. Then there exists a second-order random process X with covariance K and with trajectories in $\mathcal{H}(R)$. The conclusion holds in particular when $R \geq K$ and $\mathcal{H}(K)$ is a separable Hilbert space.*

Proof. The assumption $\mathcal{H}(K) \subseteq \mathcal{H}(R)$ implies the existence of the dominance operator L which, as a map into $\mathcal{H}(R)$, is continuous, symmetric and positive, and whose range is a subset of $\mathcal{H}(K)$ (Theorem 1.1). Since $\mathcal{H}(K)$ is a separable subset of $\mathcal{H}(R)$, L has separable range. Hence, by [16, Theorem III.2.2], L is the covariance operator of a weak second order Radon probability measure μ on $\mathcal{H}(R)$.

There exists a random element ξ in $\mathcal{H}(R)$ whose distribution $\mathbf{P}_\xi$ is μ; for example, one can take the probability space to be $(\mathcal{H}(R), \hat{\mathfrak{C}}(\mathcal{H}(R)), \mu)$ and ξ to be the identity map. The presence of weak second order implies the Pettis integrability of ξ, i.e., the existence of the mean $\mathbf{E}\xi \in \mathcal{H}(R)$, by Theorem 2.1.

Let X be the process defined by ξ, so that X has its trajectories in $\mathcal{H}(R)$. Then, according to Theorem 2.3, the mean of the process is given by $\mathbf{E}[X_t] = (\mathbf{E}\xi)(t)$, $t \in T$, and its covariance, according to (2.6), is $\langle LR_s, R_t \rangle_R$. But

$$\langle LR_s, R_t \rangle_R = \langle K_s, R_t \rangle_R = K(s,t).$$

In the case that $\mathcal{H}(K)$ is a separable Hilbert space, separability of $\mathcal{H}(K)$ as a subset of $\mathcal{H}(R)$ follows from Equation (1.4). □

Remark 6.1. One can show that a converse of the theorem holds, namely that the separability condition is necessary for the existence of the process described there.

Theorem 6.1 represents a significant strengthening of one direction of Fortet's existence result, replacing the assumption of nuclear dominance by ordinary dominance. A special case is the condition $R = K$, for which we have this result:

Corollary 6.1. *For any covariance kernel K such that the RKHS $\mathcal{H}(K)$ is separable, there exists a second-order process X with covariance K and sample paths in $\mathcal{H}(K)$.*

Corollary 6.1 is somewhat surprising since it is false for Gaussian processes with $\mathcal{H}(K)$ infinite-dimensional, as has been known for some time (see Corollary 7.1 and the remarks following it).

7. The Gaussian case

The main result of this section is that a Gaussian process with sample paths in a RKHS defines a Gaussian random element, from which nuclear dominance follows automatically (Theorem 7.1). This will suggest a second way to understand the "necessity" half of the existence result formulated by Fortet, and will lead to the generalization of Driscoll's Theorem promised in the Introduction.

We recall that a process $\{X_t, t \in T\}$ is *Gaussian* if every finite linear combination of the random variables X_t is normally distributed. A random element ξ in a Hilbert space $\mathcal{H}$ is *Gaussian* if and only if the random variable $\langle h, \xi\rangle$ is normally distributed for every $h \in \mathcal{H}$. It is trivial to see that if ξ is a Gaussian random element in the RKHS $\mathcal{H}(R)$, then the process X that it defines is Gaussian. Theorem 7.1 asserts the converse. Since a random element in $\mathcal{H}(R)$ does *not* necessarily inherit any second-order properties from the corresponding process (Section 2, Examples), it may be somewhat surprising that it *does* inherit a Gaussian distribution!

Theorem 7.1. *Let $X = \{X_t, t \in T\}$ be a Gaussian process with mean m and covariance function K. Let $\mathcal{H}(R)$ be a RKHS with $m \in \mathcal{H}(R)$. If the trajectories of X belong almost surely to $\mathcal{H}(R)$, then the random element defined by the process X is Gaussian. In particular, $R \gg K$.*

Proof. The case $m \neq 0$ can easily be reduced to the zero-mean case by considering the process $X - m$. Hence, assume $m = 0$.

Denote by ξ the random element in $\mathcal{H}(R)$ defined by the process X. We will prove that ξ is Gaussian, i.e., that $\langle \xi, f\rangle_R$ is a Gaussian random variable for every $f \in \mathcal{H}(R)$.

Clearly, for f of the form $f = \sum_{k=1}^n a_k R_{t_k}$, where $a_k \in \mathbb{R}$ and $t_k \in T$,

$$\langle \xi, f\rangle_R = \sum_{k=1}^n a_k X_{t_k},$$

which is Gaussian by assumption. The linear span V of $\{R_t,\ t \in T\}$ is dense in $\mathcal{H}(R)$ by Lemma 4.1. Hence, for every $f \in \mathcal{H}(R)$ there exists a sequence (f_n) in V such that $\|f_n - f\| \to 0$ as $n \to \infty$. Then, for every $\omega \in \Omega$,

$$\langle \xi(\omega), f_n\rangle_R \to \langle \xi(\omega), f\rangle_R \quad \text{as} \quad n \to \infty.$$

In other words, the sequence $(\langle \xi, f_n\rangle_R)$ of Gaussian random variables converges almost surely to the random variable $\langle \xi, f\rangle_R$. But then it also converges in probability, and hence its limit $\langle \xi, f\rangle_R$ is Gaussian as well [15, Lemma 1.5].

We apply Theorem 3.2, if necessary, to obtain a version η whose distribution is Radon. The covariance operator Θ of a Gauss Radon measure in a Hilbert space is nuclear by the Mourier-Prokhorov Theorem [16, Theorem IV.2.4]. But Θ is at the same time the dominance operator of R over K by Theorem 3.1. The theorem is proved. ∎

A Gaussian random element in a separable Hilbert space has strong second order, by Mourier's Theorem [14, p. 239, Theorem 2]. Thus, if we had assumed

$\mathcal{H}(R)$ is separable, then once we showed that $\langle \xi, f\rangle$ is Gaussian we could have used Corollary 3.1 to conclude that $R \gg K$.

Our first application of Theorem 7.1 is an existence result, one side of which is the Gaussian version of Theorem 6.1.

Theorem 7.2. *Let K and R be two reproducing kernels. Assume that the RKHS $\mathcal{H}(R)$ is separable.*

In order that there exist a Gaussian process with covariance K and mean $m \in \mathcal{H}(R)$ and with trajectories in $\mathcal{H}(R)$ with probability 1*, it is necessary and sufficient that $R \gg K$.*

Proof. Necessity follows from Theorem 7.1. Assume $R \gg K$. Since there exists a Gaussian process X with given mean m and covariance K [15, Proposition 3.4], it follows from Theorem 5.1 that there exists a version of X with sample paths in $\mathcal{H}(R)$. □

Remark 7.1. We may view Theorem 7.2 as one way to correct [6, Theorem 2] (see Section 1 and Remark 3.1). It is interesting to note that Piterbarg in [20] did in fact interpret Fortet's theorem in this way.

Fortet has a proof of the sufficiency part of Theorem 7.2 in his unpublished notes [7]. He utilizes the properties of the dominance operator L (our Theorem 1.1) and the given nuclearity ($R \gg K$) to infer the existence of a Gaussian measure μ on $\mathcal{H}(R)$ with covariance operator L. Here he is essentially applying the theorem of Mourier-Prokhorov [16, Theorem IV.2.4]. He concludes his proof essentially the way we proved Theorem 6.1. The present results were obtained independently [13].

Our next goal is the generalization of Driscoll's Theorem. The following result yields the "zero" part of his zero-one law.

Theorem 7.3. *Let $\{X_t, t \in T\}$ be a Gaussian process on $(\Omega, \mathcal{A})$ with mean m and covariance K. Let $\mathcal{H}(R)$ be an infinite-dimensional RKHS such that $m \in \mathcal{H}(R)$. If $R \not\gg K$, then*

$$\mathbf{P}(X_{\bullet} \in \mathcal{H}(R)) = 0,$$

where $\mathbf{P}$ denotes the completion measure on $(\Omega, \mathcal{A})$.

Proof. Let S be a countable subset of T such that $\mathcal{H}(R, S)$ is infinite-dimensional (for example, a countably infinite subset of an R-Hamel subset of T), and let Y denote the process X restricted to S. Then $X_{\bullet} \in \mathcal{H}(R,T)$ implies $Y_{\bullet} \in \mathcal{H}(R,S)$, by Proposition 4.1, so

$$\{X_{\bullet} \in \mathcal{H}(R,T)\} \subset \{Y_{\bullet} \in \mathcal{H}(R,S)\},$$

and it suffices to show that the set $\{Y_{\bullet} \in \mathcal{H}(R,S)\}$ is an event of probability zero. Thus we are reduced to proving the theorem when T is countably infinite.

In this case, enumerate T as $\{t_1, t_2, \dots\}$. Let

$$\mathcal{F} = \{S \subset T : S \text{ is finite}\};$$

then $\mathcal{F}$ is countable. Now $f \in \mathcal{H}(R)$ if and only if $f \in \mathbb{R}^T$ and Fortet's condition (4.1) holds. Clearly we may replace the supremum over $a_1, \dots, a_n \in \mathbb{R}$ by that over $a_1, \dots, a_n \in \mathbb{Q}$ (the rationals). Thus, adapting an argument of Driscoll [3], we have

$$\mathcal{H}(R) = \bigcup_{k=1}^{\infty} \bigcap_{S} \bigcap_{a_i} \{f \in \mathbb{R}^T : \frac{|\sum_i a_i f(s_i)|^2}{\sum_i \sum_j a_i a_j R(s_i, s_j)} \in [0,k]\}.$$

Both intersections are countable, and for fixed S and a_i's, the innermost set is a Borel subset of $\mathbb{R}^T$. Thus $\mathcal{H}(R)$ is measurable, and so the Kallianpur zero-one law [9] implies that $\{X_\bullet \in \mathcal{H}(R)\}$ is an event of probability zero or one. Now if the probability is one, then by Theorem 7.1 we have $R \gg K$, contradicting our assumption. Thus the probability is zero, as claimed. □

We pause to apply Theorem 7.3 when $R = K$. In this case R dominates K and the dominance map is just the identity map I. But when $\mathcal{H}(R)$ is infinite-dimensional, $\mathrm{Tr}(I) = \infty$, so that we have the following well-known result:

Corollary 7.1. *If $\{X_t, t \in T\}$ is a Gaussian process with covariance K and mean $m \in \mathcal{H}(K)$, $\mathcal{H}(K)$ infinite-dimensional, then*

$$\mathbf{P}(X_\bullet \in \mathcal{H}(K)) = 0.$$

The probability measure $\mathbf{P}$ is assumed to be complete.

Corollary 7.1 was stated by Parzen [18, Equation (34)], and proved under separability and continuity assumptions by Kallianpur [8, Theorem 5.1] and in general by LePage [10].

Combining Theorems 5.1 and 7.3, we arrive at last at the generalization of Driscoll's Theorem. We actually have two generalizations, depending on which part of Theorem 5.1 we use. Theorem 7.4 takes us as close as we can come to his zero-one law without further assumptions on the kernel R.

Theorem 7.4. *Let $\{X_t, t \in T\}$ be a Gaussian process such that $m \in \mathcal{H}(R)$. Then either*

$$\mathbf{P}(X_\bullet \in \mathcal{H}(R)) = 0, \tag{7.1}$$

or there exists a version Y of X such that

$$\mathbf{P}(Y_\bullet \in \mathcal{H}(R)) = 1, \tag{7.2}$$

depending on whether $R \not\gg K$ or $R \gg K$, respectively. $\mathbf{P}$ is assumed to be complete.

Theorem 7.5. *Let $\{X_t, t \in T\}$ be a Gaussian process such that $m \in \mathcal{H}(R)$. Assume that d_R is a metric, that $\mathcal{H}(R)$ is separable, and that the trajectories of X are d_R-continuous functions on T. Then*

$$\mathbf{P}(X_\bullet \in \mathcal{H}(R)) = 0 \text{ or } 1,$$

and equals 1 if and only if $R \gg K$. $\mathbf{P}$ is assumed to be complete.

8. Concluding remarks

As mentioned in the Introduction, Driscoll applied his theorem to find the Bayes estimator of a mean function m under a Gaussian prior, using the norm of $\mathcal{H}(R)$ to define a quadratic loss function. This problem may be viewed as a prediction problem for a random signal in $\mathcal{H}(R)$. Having generalized his theorem, we are in a position to solve the corresponding prediction problem in a more general setting. This will appear in a future paper.

Theorem 7.1 shows that a Gaussian process with sample paths in a RKHS defines a Gaussian random element in that space. It is well known that the even-order moments of a zero-mean, normally distributed random variable may be expressed explicitly in terms of its variance. It turns out that for a zero-mean Gaussian random element ξ in a RKHS one can similarly compute the even-order moments of $\|\xi\|$ in terms of the trace of the covariance operator. This will also be described elsewhere.

In the present paper we have been able to remove Driscoll's assumption that T be a separable metric space and that the kernel R be continuous by introducing a "natural" metric on T defined solely by the kernel R. From the results of Section 5 it is natural to ask whether one may measure the probability that a given process X has trajectories which are d_R-continuous. The answer is unknown at this time.

It would also be of interest to know whether a process X with covariance K and sample paths in $\mathcal{H}(R)$ automatically has its mean function in $\mathcal{H}(R)$ when $R \geq K$. Under these conditions we may certainly define a random element ξ in $\mathcal{H}(R)$, but it is not known, for example, whether ξ has weak first order. A positive answer to this question would allow us to strengthen Theorem 3.1 by deleting the assumption that the mean of X is in $\mathcal{H}(R)$.

9. Acknowledgments

We would like to thank the editor for his expeditious handling of this paper, and the referee for an insightful and helpful review. We would also like to thank Professor Bernard Bru for kindly providing Fortet's unpublished notes [7] to us.

A question raised by Professor Thomas Kurtz during a UW-Madison Probability Seminar on this material led to Corollary 6.1.

References

[1] N. Aronszajn, *Theory of reproducing kernels*, Transactions of the American Mathematical Society **68** (1950), 337–404. MR **14:**479c

[2] Michael F. Driscoll, *Estimation of the mean value function of a Gaussian process*, Ph.D. thesis, University of Arizona, 1971.

[3] ———, *The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **26** (1973), 309–316. MR **51:**6949

[4] ———, *The signal-noise problem – a solution for the case that signal and noise are Gaussian and independent*, Journal of Applied Probability **12** (1974), 183–187. MR **51:**2222

[5] Robert M. Fortet, *Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires*, Annales de l'Institut Henri Poincaré, B **IX** (1973), 41–58. MR **49:**9932

[6] ———, *Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, A **278** (1974), 1439–1440. MR **51:**6950

[7] ———, *Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires*, unpublished notes on [6], 1974.

[8] Gopinath Kallianpur, *The role of reproducing kernel Hilbert spaces in the study of Gaussian processes*, Advances in Probability and Related Topics, vol. 2, Marcel Dekker, 1970, pp. 49–83. MR **44:**1096

[9] ———, *Zero-one laws for Gaussian processes*, Transactions of the American Mathematical Society **149** (1970), 199–211. MR **42:**1200

[10] Raoul D. LePage, *Subgroups of paths and reproducing kernels*, The Annals of Probability **1** (1973), 345–347. MR **50:**3327

[11] M. Loève, *Fonctions aléatoires du second ordre*, Supplement to P. Lévy, Processus Stochastiques et Mouvement Brownien, Gauthier-Villars, Paris, 1948.

[12] Eugene Lukacs, *Stochastic Convergence*, Probability and Mathematical Statistics, vol. 30, Academic Press, New York, 1975. MR **51:**11599

[13] Milan N. Lukić, *Stochastic processes having sample paths in reproducing kernel Hilbert spaces with an application to white noise analysis*, Ph.D. thesis, University of Wisconsin, Milwaukee, 1996.

[14] Edith Mourier, *Eléments aléatoires dans un espace de Banach*, Annales de l'Institut Henri Poincaré **13** (1953), 161–244. MR **16:**268a

[15] Jacques Neveu, *Processus Aléatoires Gaussiens*, Publications du Séminaire de Mathématiques Supérieures, Les Presses de l'Université de Montréal, 1968. MR **42:**6923

[16] Н. Н. Вахания, В. И. Тариеладзе и С. А. Чобанян, *Вероятностные распределения в банаховых пространствах*, Nauka, Moscow, 1985, English translation [22].

[17] Emanuel Parzen, *Statistical inference on time series by Hilbert space methods I*, Tech. Report 23, Statistics Department, Stanford University, 1959, reprinted in [19], Chapter 13.

[18] ———, *Probability density functionals and reproducing kernel Hilbert spaces*, Proceedings of the Symposium on Time Series Analysis (New York) (M. Rosenblatt, ed.), John Wiley & Sons, Inc., 1963, reprinted in [19], Chapter 17. MR **26:**7119

[19] ———, *Time Series Analysis Papers*, Holden-Day, San Francisco, 1967. MR **36:**1069

[20] V. Piterbarg, *Review 1b168 of* [6], Referativnyi Zhurnal Matematika **1**–**2** (1975), 25 (Russian), Izdatelstvo Akademii Nauk SSSR.

[21] M. Talagrand, *Regularity of Gaussian processes*, Acta Mathematica **159** (1987), 99–149. MR **89b:**60106

[22] N. N. Vakhaniya, V. I. Tarieladze and S. A. Chobanyan, *Probability Distributions on Banach Spaces*, Riedel, Dordrecht, Holland, 1987, translated by W. A. Woyczynski. MR **97k:**60007

[23] N. N. Vakhania and V. I.Tarieladze, *Covariance operators of probability measures in locally convex spaces*, Theory of Probability and its Applications **23** (1978), 1–21. MR **58:**24442

Department of Mathematics, Viterbo University, 815 South 9th Street, La Crosse, Wisconsin 54601

E-mail address: mnlukic@viterbo.edu

URL: http://www.viterbo.edu/personalpages/faculty/MLukic/index.htm

Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413

E-mail address: beder@uwm.edu

URL: http://www.uwm.edu/~beder/