Small Sample Spaces for Gaussian Processes

TONI KARVONEN¹,²,*

¹The Alan Turing Institute, 96 Euston Road, London NW1 2DB, United Kingdom
²University of Helsinki, Department of Mathematics and Statistics, Pietari Kalmin katu 5, 00560 Helsinki, Finland
E-mail: *toni.karvonen@helsinki.fi

It is known that the membership in a given reproducing kernel Hilbert space (RKHS) of the samples of a Gaussian process \( X \) is controlled by a certain nuclear dominance condition. However, it is less clear how to identify a “small” set of functions (not necessarily a vector space) that contains the samples. This article presents a general approach for identifying such sets. We use scaled RKHSs, which can be viewed as a generalisation of Hilbert scales, to define the sample support set as the largest set which is contained in every element of full measure under the law of \( X \) in the \( \sigma \)-algebra induced by the collection of scaled RKHS. This potentially non-measurable set is then shown to consist of those functions that can be expanded in terms of an orthonormal basis of the RKHS of the covariance kernel of \( X \) and have their squared basis coefficients bounded away from zero and infinity, a result suggested by the Karhunen–Loève theorem.

Keywords: Gaussian processes; sample path properties; reproducing kernel Hilbert spaces

1. Introduction

Let \( K : T \times T \to \mathbb{R} \) be a positive-semidefinite kernel on a set \( T \) and consider any Gaussian process \((X(t))_{t \in T}\) with mean zero and covariance \( K \), which we denote \((X(t))_{t \in T} \sim \mathcal{GP}(0, K)\). Let \( H(K) \) be the reproducing kernel Hilbert space (RKHS) of \( K \) equipped with inner product \( \langle \cdot , \cdot \rangle_K \) and norm \( \| \cdot \|_K \).

It is a well-known fact, apparently originating with Parzen [28], that the samples of \( X \) are not contained in \( H(K) \) if this space is infinite-dimensional. Furthermore, Driscoll [9]; Fortet [12]; and Lukić and Beder [24] have used the zero-one law of Kallianpur [15] to show essentially that, given another kernel \( R \) and under certain mild assumptions,

\[
P[X \in H(R)] = 1 \text{ if } R \gg K \text{ and } P[X \in H(R)] = 0 \text{ if } R \not\gg K,
\]

where \( R \gg K \) signifies that \( R \) dominates \( K \) (i.e., \( H(K) \subset H(R) \)) and, moreover, that the dominance is nuclear (see Section 4.1 for details). This Driscoll’s theorem is an exhaustive tool for verifying whether or not the samples from a Gaussian process are contained in a given RKHS. A review of the topic can be found in [13, Chapter 4]. Two questions now arise:

- How to construct a kernel \( R \) such that \( R \gg K \)?
- Is it possible to exploit the fact that \( P[X \in H(R_1) \setminus H(R_2)] = 1 \) for any kernels such that \( R_1 \gg K \) and \( R_2 \not\gg K \) to identify in some sense the smallest set of functions which contains the samples with probability one?

Answers to questions such as these are instructive for theory and design of Gaussian process based learning [40, 11], emulation and approximation [41, 18], and optimisation [3] methods. For simplicity we assume that the domain \( T \) is a complete separable metric space, that the kernel \( K \) is continuous and its RKHS is separable, and that the samples of \( X \) are continuous. Although occasionally termed “rather restrictive” [9, p. 309], these continuity assumptions are satisfied by the vast majority of domains and Gaussian processes commonly used in statistics and machine learning literature [35, 31], such as
stationary processes with Gaussian or Matérn covariance kernels. Our motivation for imposing these restrictions is that they imply that RKHSs are measurable.

1.1. Contributions

First, we present a flexible construction for a kernel $K$ such that $R \gg K$. For any orthonormal basis $\Phi = (\phi_n)_{n=1}^\infty$ of $H(K)$ the kernel can be written as $K(t,t') = \sum_{n=1}^\infty \phi_n(t)\phi_n(t')$ for all $t,t' \in T$. Given a positive sequence $A = (\alpha_n)_{n=1}^\infty$ such that $\sum_{n=1}^\infty \alpha_n \phi_n(t)^2 < \infty$ for all $t \in T$, we define the scaled kernel

$$K_{A,\Phi}(t,t') = \sum_{n=1}^\infty \alpha_n \phi_n(t)\phi_n(t').$$

This is a significant generalisation of the concept of powers of kernels which has been previously used to construct RKHSs which contain the samples by Steinwart [36]. The sequence $A$ is called a $\Phi$-scaling of $H(K)$. If $\alpha_n \to \infty$ as $n \to \infty$, the corresponding scaled RKHS, $H(K_{A,\Phi})$, is a proper superset of $H(K)$, though not necessarily large enough to contain the samples of $X$. We show that convergence of the series $\sum_{n=1}^\infty \alpha_n^{-1}$ controls whether or not samples are contained in $H(K_{A,\Phi})$. If $\sum_{n=1}^\infty \alpha_n^{-1}$ converges slowly, we can therefore interpret $H(K_{A,\Phi})$ as a “small” RKHS which contains the samples.

Main Result I (Theorem 4.3). Let $\Phi = (\phi_n)_{n=1}^\infty$ be an orthonormal basis of $H(K)$ and $A = (\alpha_n)_{n=1}^\infty$ a $\Phi$-scaling of $H(K)$. If $K_{A,\Phi}$ is continuous and

$$d_K(t,t') = \|K(\cdot,t) - K(\cdot,t')\|_K = \sqrt{K(t,t) - 2K(t,t') + K(t',t')}$$

is a metric on $T$, then either

$$\mathbb{P}[X \in H(K_{A,\Phi})] = 0 \quad \text{and} \quad \sum_{n=1}^\infty \frac{1}{\alpha_n} = \infty$$

or

$$\mathbb{P}[X \in H(K_{A,\Phi})] = 1 \quad \text{and} \quad \sum_{n=1}^\infty \frac{1}{\alpha_n} < \infty.$$
Main Result II (Theorems 4.8 and 4.11). Let $\Phi = (\phi_n)_{n=1}^\infty$ be an orthonormal basis of $H(K)$ and suppose there is a $\Phi$-scaling $A = (\alpha_n)_{n=1}^\infty$ such that $\sum_{n=1}^\infty \alpha_n^{-1} < \infty$ and $K_{\Phi}$ is continuous. Let $\mathcal{S}(\mathbb{R})$ be the $\sigma$-algebra generated by the collection of scaled RKHSs consisting of continuous functions and $\mathcal{S}_H(K)$ the largest subset of $C(T)$ that is contained in every $H \in \mathcal{S}(\mathbb{R})$ such that $\mathbb{P}[X \in H] = 1$. Suppose that $d_H(t, t') = \|K(\cdot, t) - K(\cdot, t')\|_K$ is a metric on $T$. Then $\mathcal{S}_H(K)$ consists precisely of the functions $f = \sum_{n=1}^\infty f_n \phi_n$ such that

$$\liminf_{n \to \infty} f_n^2 > 0 \quad \text{and} \quad \sup_{n \geq 1} f_n^2 < \infty. \quad (1.2)$$

Furthermore, for every $H \in \mathcal{S}(\mathbb{R})$ such that $\mathbb{P}[X \in H] = 1$ there exists $F \in \mathcal{S}(\mathbb{R})$ such that $\mathcal{S}_H(K)$ is a proper subset of $F$ and $F$ is a proper subset of $H$.

The set $\mathcal{S}_H(K)$ may fail to be measurable. The latter part of the above result is therefore important in demonstrating that it is possible to construct, using countably many elementary set operations of scaled RKHSs, sets which are arbitrarily close to $\mathcal{S}_H(K)$ and contain the samples. At its core this is a manifestation of the classical result that there does not exist a meaningful notion of a boundary between convergent and divergent series [19, § 41]. The general form of the Karhunen–Loève theorem is useful in explaining the characterisation in (1.2). If $(\phi_n)_{n=1}^\infty$ is any orthonormal basis of $H(K)$, then the Gaussian process can be written as $X(t) = \sum_{n=1}^\infty \zeta_n \phi_n(t)$ for all $t \in T$, where $\zeta_n$ are independent standard normal random variables. In general the expansion converges in $L^2(\mathbb{P})$, but if almost all samples of the process are continuous, then convergence is also uniform on $T$ with probability one [1, Theorem 3.8]. Because $\|X(t)\|^2_K = \sum_{n=1}^\infty \zeta_n^2$ and $\mathbb{P}[\zeta_n^2] = 1$ for every $n$, the Karhunen–Loève expansion suggests, somewhat informally speaking, that the samples can be found in the set of functions $f = \sum_{n=1}^\infty f_n \phi_n$ for which the sequence $(f_n^2)_{n=1}^\infty$ “is of constant order”.

1.2. On Measurability and Continuity

Suppose for a moment that $T$ is an arbitrary uncountable set, $K$ a positive-semidefinite kernel on $T$ such that its RKHS $H(K)$ is infinite-dimensional, and $(X(t))_{t \in T} \sim \mathcal{G}(0, K)$ a generic Gaussian process defined on a generic probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\mathbb{R}^T$ be the collection of real-valued functions on $T$ and $\mathcal{B}$ the $\sigma$-algebra generated by cylinder sets of the form $\{f \in \mathbb{R}^T : (f(t_1), \ldots, f(t_n)) \in B^n\}$ for any $n \in \mathbb{N}$ and any Borel set $B^n \subset \mathbb{R}^n$. Let $\Phi_X(\omega) = X(\cdot, \omega)$. Then $\mu_X = \mathbb{P} \circ \Phi_X^{-1}$ is the law of $X$ on the measurable space $(\mathbb{R}^T, \mathcal{B})$. Consequently,

$$\mathbb{P}(X \in H) = \hat{\mu}(\{\omega \in \Omega : X(\cdot, \omega) \in H\})$$

for $H \in \mathcal{B}$. Let $(\mathbb{R}^T, \mathcal{B}_0, \hat{\mu}_{X,0})$ be the completion of $(\mathbb{R}^T, \mathcal{B}, \hat{\mu}_X)$ and $R : T \times T \to \mathbb{R}$ a positive-semidefinite kernel and $H(R)$ its RKHS. The following facts are known about measurability of $H(K)$ and $H(R)$:

- In general, $H(R) \notin \mathcal{B}$. For example, if $T$ is equipped with a topology and $R$ is continuous on $T \times T$, then $H(R) \subset C(T)$ has proved, however, no non-empty subset of $C(T)$ can be an element of $\mathcal{B}$.
- LePage [21, p. 347] has proved that $H(K) \in \mathcal{B}_0$ and $\hat{\mu}_{X,0}(H(K)) = 0$, a claim which originates with Parzen [27, 28]. A less general version which requires separability and continuity has been proved by Kallianpur [16].
- More generally, if the RKHS $H(R)$ is infinite-dimensional and $R \not \sim K$, then $H(R) \in \mathcal{B}_0$ and $\hat{\mu}_{X,0}(H(R)) = 0$. This result is contained in the proof of Theorem 7.3 in [24]. See also [13, Proposition 4.5.1].
• LePage [21, Corollary 2] has proved a dichotomy result which states that if \( G \subset \mathbb{R}^T \) is an additive group and \( G \in \mathcal{B}_0 \), then either \( \mu_{X,0}(G) = 0 \) or \( \mu_{X,0}(G) = 1 \). Furthermore \( \mu_{X,0}(G) = 1 \) implies that \( H(K) \subset G \). This is the most general version of the famous zero-one law of Kallianpur [15, Theorem 2].

It appears that not much more can be said without imposing additional structure or constructing versions of \( X \), as is done in [24].

Suppose that \((T, d_T)\) is a complete separable metric space, kernel \( K \) is continuous, and almost all samples of \((X(t))_{t \in T} \sim \mathcal{G}\mathcal{P}(0, K)\) are continuous (i.e., the \((\mathcal{B}_0, \mu_0)\)-outer measure of \( C(T) \) is one). Define the probability space \((C(T), \mathcal{B}, \mu_X)\) as

\[
\mathcal{B} = C(T) \cap \mathcal{B}_0 \quad \text{and} \quad \mu_X(C(T) \cap H) = \mu_{X,0}(H) \quad \text{for} \quad H \in \mathcal{B}_0.
\]

The rest of this article is concerned with \((C(T), \mathcal{B}, \mu_X)\) and it is to be understood that \( \mathbb{P}(X \in H) \) stands for \( \mu_X(H) \) for any \( H \in \mathcal{B} \). In this setting Driscoll [9, p. 313] has proved that \( H(R) \in \mathcal{B} \) if \( R \) is continuous and positive-definite. By using Theorem 1.1 in [12] (Theorem 4.1 in [24]) one can generalise this result to the case where \( R \) is continuous and positive-semidefinite; see the proof of Theorem 7.3 in [24].

### 1.3. Notation and Terminology

For non-negative real sequences \((a_n)_{n=1}^\infty\) and \((b_n)_{n=1}^\infty\) we write \(a_n \leq b_n\) if there is \( C > 0\) such that \(a_n \leq Cb_n\) for all sufficiently large \(n\). If we have both \(a_n \leq b_n\) and \(b_n \leq a_n\), we write \(a_n \asymp b_n\). If \(a_n/b_n \to 1\) as \(n \to \infty\), we write \(a_n \sim b_n\). For two sets \(F\) and \(G\) we use \(F \subseteq G\) to indicate that \(F\) is a proper subset of \(G\). A kernel \(R \colon T \times T \to \mathbb{R}\) is positive-semidefinite if

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j R(t_i, t_j) \geq 0
\]

for any \(N \geq 1\), \(a_1, \ldots, a_N \in \mathbb{R}\) and \(t_1, \ldots, t_N \in T\). In the remainder of this article positive-semidefinite kernels are simply referred to as kernels. If the inequality in (1.4) is strict for any pairwise distinct \(t_1, \ldots, t_N\), the kernel is said to be positive-definite.

### 1.4. Standing Assumptions

For ease of reference our standing assumptions are collected here. We assume that (i) \((T, d_T)\) is a complete separable metric space; (ii) the covariance kernel \(K \colon T \times T \to \mathbb{R}\) is continuous and positive-semidefinite on \(T \times T\); (iii) the RKHS \(H(K)\) induced by \(K\) is infinite-dimensional and separable\(^2\); and (iv) \((X(t))_{t \in T} \sim \mathcal{G}\mathcal{P}(0, K)\) is a zero-mean Gaussian process on a probability space \((\Omega, \mathcal{A}, \mathbb{P})\) with continuous paths. The law \(\mu_X\) of \(X\) is defined on the measurable space \((C(T), \mathcal{B})\) which was constructed in Section 1.2. Some of our results have natural generalisations for general second-order stochastic processes; see [24], in particular Sections 2 and 5, and [36]. These generalisations are not pursued here.

\(^2\)Most famously, separable RKHSs are induced by Mercer kernels, which are continuous kernels defined on compact subsets of \(\mathbb{R}^d\) [29, Section 11.3].
2. Related Work

Reproducing kernel Hilbert spaces which contain the samples of \((X(t))_{t \in T} \sim \mathcal{GP}(0, K)\) have been constructed by means of integrated kernels in \cite{23}, convolution kernels in \cite{5} and \cite[Section 3.1]{11}, and, most importantly, powers of RKHSs \cite{39} in \cite{36} and \cite[Section 4]{17}. Namely, let \(T\) be a compact metric space, \(K\) a continuous kernel on \(T \times T\), and \(\nu\) a finite and strictly positive Borel measure on \(T\). Then the integral operator \(T_\nu\), defined for \(f \in L^2(\nu)\) via

\[
(T_\nu f)(t) = \int_T K(t, t') f(t') \, d\nu(t'),
\]

has decreasing and positive eigenvalues \((\lambda_n)_{n=1}^\infty\), which vanish as \(n \to \infty\), and eigenfunctions \((\psi_n)_{n=1}^\infty \subset H(K)\) such that \(\sqrt{\lambda_n} \psi_n\) is an orthonormal basis of \(H(K)\). The kernel has the uniformly convergent Mercer expansion \(K(t, t') = \sum_{n=1}^\infty \lambda_n \psi_n(t) \psi_n(t')\) for all \(t, t' \in T\). For \(\theta > 0\), the \(\theta\)th power of \(K\) is defined as

\[
K^{(\theta)}(t, t') = \sum_{n=1}^\infty \lambda_n^\theta \psi_n(t) \psi_n(t').
\]

The series (2.2) converges if \(\sum_{n=1}^\infty \lambda_n^\theta \psi_n(t)^2 < \infty\) for all \(t \in T\). Furthermore, if \(\theta_1 < \theta_2\), then \(H(K^{(\theta_2)}) \subset H(K^{(\theta_1)})\). One can prove that \(\mathbb{P}[X \in H(K^{(\theta)})] = 1\) if and only if \(\sum_{n=1}^\infty \lambda_n^{1-\theta} < \infty\) \cite[Theorem 5.2]{36}. When it comes to sample properties, the power kernel construction has two significant downsides: (i) The measure \(\nu\) is a nuisance parameter. If one is only interested in sample path properties of Gaussian processes this measure should not have an intrinsic part to play in the analysis and results. (ii) The construction is somewhat inflexible and unsuitable for infinitely smooth kernels. Because the power RKHS \(H(K^{(\theta)})\) consists precisely of the functions

\[
f = \sum_{n=1}^\infty f_n \lambda_n^{1/2} \psi_n\]

such that \(\sum_{n=1}^\infty f_n^2 \lambda_n^{1-\theta} < \infty\)

and \(\lambda_n \to 0\) as \(n \to \infty\), how much larger \(H(K^{(\theta)})\) is than \(H(K)\), given by \(\theta = 1\), is determined by rate of decay of the eigenvalues. Power RKHSs are more descriptive and fine-grained when the kernel is finitely smooth and its eigenvalues have a polynomial decay (e.g., Matérn kernels) than when the kernel is infinitely smooth with exponential eigenvalue decay (e.g., Gaussian). The change from a polynomial decay condition \(\sum_{n=1}^\infty f_n^2 n^{-\alpha}\) for the coefficients \((f_n)_{n=1}^\infty\) to \(\sum_{n=1}^\infty f_n^2 n^{-(1-\theta)} < \infty\) is arguably less substantial than that from an exponential decay condition \(\sum_{n=1}^\infty f_n^2 e^{-bn} < \infty\) to \(\sum_{n=1}^\infty f_n^2 e^{-b(1-\theta)n} < \infty\). Indeed, as pointed out by Kanagawa et al. \cite[Section 4.4]{17}, when the kernel is Gaussian every \(H(K^{(\theta)})\) with \(\theta < 1\) contains the samples with probability one, which renders powers of RKHSs of dubious utility in that setting. The relationship between powers of RKHSs and scaled RKHSs is discussed more at the end of Section 3. In Section 5 we demonstrate that scaled RKHSs are more useful in describing sample path properties of Gaussian processes defined by infinitely smooth kernels than powers of RKHSs.

To the best of our knowledge, the question about a “minimal” set which contains the samples with probability one has received only cursory discussion in the literature. Perhaps the most relevant digression on the topic is an observation by Steinwart \cite[pp. 369–370]{36}, given here in a somewhat
applied form and without some technicalities, that the samples are contained in the set
\[ W^s_2(T) \setminus W^s_2(T) \]  
with probability one if \( H(K) \) is norm-equivalent to the fractional Sobolev space \( W^{s+d/2}_2(T) \) for \( s > 0 \) on a suitable domain \( T \subset \mathbb{R}^d \). In the Sobolev case the samples are therefore “\( d/2 \) less smooth” than functions in the RKHS of \( K \). Because \( W^s_2(T) = \bigcup_{r \geq s} W^r_2(T) \), the set in (2.3) has the same form as the sample support set in (4.2). This observation is, of course, a general version of the familiar result that the sample paths of the Brownian motion, whose covariance kernel \( K(t, t') = \min\{t, t'\} \) on \( T = [0, 1] \) induces the Sobolev space \( W^s_2([0, 1]) \) with zero boundary condition at the origin, have regularity \( 1/2 \) in a sense that they are almost surely \( \alpha \)-Hölder continuous if and only if \( \alpha < 1/2 \). That is, there is \( C > 0 \) such that, for almost every \( \omega \in \Omega \),
\[ |X(t, \omega) - X(t', \omega)| \leq C |t - t'|^{\alpha} \]
for all \( t, t' \in [0, 1] \) and any \( \alpha < 1/2 \). However, Lévy’s modulus of continuity theorem [26, Section 1.2] improves this to
\[ |X(t, \omega) - X(t', \omega)| \leq C \sqrt{\log(1/h)} \]
when \( h = |t - t'| \) is sufficiently small. Since the Sobolev space \( W^s_2(T) \) consists of those functions \( f : T \to \mathbb{R} \) which admit an \( L^2(\mathbb{R}^d) \)-extension \( f_e : \mathbb{R}^d \to \mathbb{R} \) whose Fourier transform satisfies
\[ \int_{\mathbb{R}^d} \big(1 + \|\xi\|^2\big)^s |\hat{f}_e(\xi)|^2 \, d\xi < \infty, \]  
Lévy’s modulus of continuity theorem suggests replacing the weight in (2.4) with, for example, \( (1 + \|\xi\|^2)^s \log(1 + \|\xi\|) \) so that the resulting function space is a proper superset of \( W^s_2(T) \) and a proper subset of \( W^r_2(T) \) for every \( r < s \) and hence a proper subset of the set in (2.3). Some results and discussion in [34] and [18, Section 4.2] have this flavour.

Finally, we remark that classical results about the topological support of a Gaussian measure are distinct from the results in this article. Let \( C(T) \) be equipped with the standard supremum norm \( \| \cdot \|_\infty \). The topological support, \( \text{supp}_{C(T)}(\mu_X) \), of the measure \( \mu_X \) is the set
\[ \text{supp}_{C(T)}(\mu_X) = \{ f \in C(T) : \mu_X(B(f, r)) > 0 \text{ for all } r > 0 \}, \]
where \( B(f, r) \) is the \( f \)-centered \( r \)-ball in \((C(T), \| \cdot \|_\infty)\). It is a classical result that [16, Theorem 3]
\[ \text{supp}_{C(T)}(\mu_X) = \overline{H(K)}, \]
where \( \overline{H(K)} \) is the closure of \( H(K) \) in \((C(T), \| \cdot \|_\infty)\). In other words, the topological support of \( \mu_X \) contains every continuous function \( f \) such that for every \( \varepsilon > 0 \) there exist \( g_\varepsilon \in H(K) \) satisfying \( \|f - g_\varepsilon\|_\infty < \varepsilon \). Now, recall that a kernel \( R \) is universal if \( H(R) \) is dense in \( C(T) \) [37, Section 4.6]. Most kernels of interest to practitioners are universal, including Gaussians, Matérns, and power series kernels. But, by definition, the closure of the RKHS of a universal kernel equals \( C(T) \). Therefore
\[ \text{supp}_{C(T)}(\mu_X) = \overline{H(K)} = C(T) \]
if \( K \) is a universal kernel. This result does not provide any information about the samples because we have assumed that the samples are continuous to begin with. See, for example, Section 3.6 in [2] for further results on general topological supports of Gaussian measures.
3. Scaled Reproducing Kernel Hilbert Spaces

Let $\Phi = (\phi_n)_{n=1}^{\infty}$ be any orthonormal basis of $H(K)$. Then the kernel has the pointwise convergent expansion

$$K(t, t') = \sum_{n=1}^{\infty} \phi_n(t) \phi_n(t') \quad \text{for all} \quad t, t' \in T$$

and, by the standard characterisation of a separable Hilbert space, the RKHS consists of precisely those functions $f : T \to \mathbb{R}$ that admit an expansion

$$f = \sum_{n=1}^{\infty} f_n \phi_n \quad \text{for coefficients such that} \quad \sum_{n=1}^{\infty} f_n^2 < \infty.$$ 

By the Cauchy–Schwarz inequality this expansion converges pointwise on $T$. For given functions $f = \sum_{n=1}^{\infty} f_n \phi_n$ and $g = \sum_{n=1}^{\infty} g_n \phi_n$ in the RKHS the inner product is

$$\langle f, g \rangle_{H(K)} = \sum_{n=1}^{\infty} f_n g_n.$$ 

**Definition 3.1 (Scaled kernel and RKHS).** We say that a positive sequence $A = (\alpha_n)_{n=1}^{\infty}$ is a $\Phi$-scaling of $H(K)$ if

$$\sum_{n=1}^{\infty} \alpha_n \phi_n(t)^2 < \infty \quad \text{for every} \quad t \in T. \quad (3.1)$$

The kernel

$$K_{A, \Phi}(t, t') = \sum_{n=1}^{\infty} \alpha_n \phi_n(t) \phi_n(t') \quad (3.2)$$

is called a scaled kernel and its RKHS $H(K_{A, \Phi})$ a scaled RKHS.

See [30, 42, 43] for prior appearances of scaled kernels under different names and not in the context of Gaussian processes. Although many of the results in this section have appeared in some form in the literature, all proofs are included here for completeness.

**Proposition 3.2.** Let $\Phi = (\phi_n)_{n=1}^{\infty}$ be an orthonormal basis of $H(K)$ and $A = (\alpha_n)_{n=1}^{\infty}$ a $\Phi$-scaling of $H(K)$. Then (i) the scaled kernel $K_{A, \Phi}$ is positive-semidefinite, (ii) the collection $(\sqrt{\alpha_n} \phi_n)_{n=1}^{\infty}$ is an orthonormal basis of $H(K_{A, \Phi})$, and (iii) the scaled RKHS is

$$H(K_{A, \Phi}) = \left\{ f = \sum_{n=1}^{\infty} f_n \sqrt{\alpha_n} \phi_n : \|f\|_{K_{A, \Phi}}^2 = \sum_{n=1}^{\infty} f_n^2 \alpha_n < \infty \right\} \quad (3.3)$$

where convergence is pointwise, and its inner product is

$$\langle f, g \rangle_{K_{A, \Phi}} = \sum_{n=1}^{\infty} \frac{f_n g_n}{\alpha_n} \quad (3.4)$$

for any $f = \sum_{n=1}^{\infty} f_n \phi_n$ and $g = \sum_{n=1}^{\infty} g_n \phi_n$ in $H(K_{A, \Phi})$. 

Proof. By the Cauchy–Schwarz inequality and (3.1),
\[
\sum_{n=1}^{\infty} |\alpha_n \phi_n(t) \phi_n(t')| \leq \left( \sum_{n=1}^{\infty} \alpha_n \phi_n(t)^2 \right)^{1/2} \left( \sum_{n=1}^{\infty} \alpha_n \phi_n(t')^2 \right)^{1/2}
\]
for any \( t, t' \in T \). This proves that the scaled kernel in (3.2) is well-defined via an absolutely convergent series. To verify that \( K_{A,\Phi} \) is positive-semidefinite, note that, for any \( N \geq 1, a_1, \ldots, a_N \in \mathbb{R} \), and \( t_1, \ldots, t_N \in T \),
\[
\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j K_{A,\Phi}(t_i, t_j) = \sum_{n=1}^{\infty} \alpha_n \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \phi_n(t_i) \phi_n(t_j) = \sum_{n=1}^{\infty} \alpha_n \left( \sum_{i=1}^{N} a_i \phi_n(t_i) \right)^2
\]
is non-negative because each \( \alpha_n \) is positive. Because
\[
\sum_{n=1}^{\infty} |f_n \sqrt{\alpha_n \phi_n(t)}| \leq \left( \sum_{n=1}^{\infty} f_n^2 \right)^{1/2} \left( \sum_{n=1}^{\infty} \alpha_n \phi_n(t)^2 \right)^{1/2} < \infty \quad \text{for every} \quad t \in T
\]
if \( \sum_{n=1}^{\infty} f_n^2 < \infty \), the space defined in (3.3) and (3.4) is a Hilbert space of functions with an orthonormal basis \( \{ \sqrt{\alpha_n \phi_n} \}_{n=1}^{\infty} \). Since \( K_{A,\Phi}(t, t') = \sum_{n=1}^{\infty} \alpha_n \phi_n(t) \phi_n(t') \), the scaled kernel is the unique reproducing kernel of this space [e.g., 25, Theorem 9].

A scaled RKHS depends on the ordering of the orthonormal basis of \( H(K) \) used in its construction. For example, let \( \Phi = (\phi_n)_{n=1}^{\infty} \) be an orthonormal basis of \( H(K) \) and suppose that the coefficients \( \alpha_n = n \) define a \( \Phi \)-scaling of \( H(K) \). Define another ordered orthonormal basis \( \Psi = (\psi_n)_{n=1}^{\infty} \) by setting \( \psi_{2n+1} = \phi_{2n} \) for \( n \geq 0 \) and interleaving the remaining \( \phi_n \) to produce
\[
\Psi = (\phi_1, \phi_3, \phi_2, \phi_5, \phi_4, \phi_6, \phi_8, \phi_7, \phi_{16}, \ldots).
\]
Now, the function \( f = \sum_{n=0}^{\infty} \phi_{2n} = \sum_{n=1}^{\infty} f_{\psi_n, \psi_n} \phi_n \) is in \( H(K_{A,\Psi}) \) because
\[
\| f \|^2_{K_{A,\Psi}} = \sum_{n=1}^{\infty} f_{\psi_n, \psi_n}^2 \alpha_n = \sum_{n=0}^{\infty} \frac{1}{2^n} < \infty
\]
buts not in \( H(K_{A,\Phi}) \) because \( f = \sum_{n=0}^{\infty} \phi_{2n} = \sum_{n=0}^{\infty} \psi_{2n+1} = \sum_{n=1}^{\infty} f_{\psi_n, \psi_n} \) and therefore
\[
\| f \|^2_{K_{A,\Phi}} = \sum_{n=1}^{\infty} f_{\psi_n, \psi_n}^2 \alpha_n = \sum_{n=0}^{\infty} \frac{1}{2n + 1} = \infty.
\]
In practice, there is usually a natural ordering of the orthonormal basis. For instance, the decreasing eigenvalues specify a natural ordering for a basis obtained from Mercer’s theorem (see Section 2) or the basis may have a polynomial factor, the degree of which specifies an ordering (see the kernels in Sections 5.2 and 5.3).

The following results compare sizes of scaled RKHSs: the faster \( \alpha_n \) grows, the larger the RKHS \( H(K_{A,\Phi}) \) is. A number of additional properties between scaled RKHSs can be proved in a similar manner but are not needed in the developments of this article. Some of the below results or their variants can be found in the literature. In particular, see [42, Section 6] and [43, Section 4] for a version of Proposition 3.5 and some additional results.
Proposition 3.3. Let \( \Phi = (\phi_n)_{n=1}^{\infty} \) be an orthonormal basis of \( H(K) \) and \( A = (\alpha_n)_{n=1}^{\infty} \) and \( B = (\beta_n)_{n=1}^{\infty} \) two \( \Phi \)-scalings of \( H(K) \). Then
\[
H(K_{B,\Phi}) \subset H(K_{A,\Phi}) \quad \text{if and only if} \quad \beta_n \leq \alpha_n.
\]
In particular,
\[
H(K) \subset H(K_{A,\Phi}) \quad \text{if and only if} \quad \inf_{n \geq 1} \alpha_n > 0.
\]

Proof. If \( \beta_n \leq \alpha_n \), then for any \( f = \sum_{n=1}^{\infty} f_n \phi_n \in H(K_{B,\Phi}) \) we have
\[
\| f \|_{K_{A,\Phi}}^2 = \sum_{n=1}^{\infty} \frac{f_n^2}{\alpha_n} = \sum_{n=1}^{\infty} \frac{\beta_n}{\alpha_n} \frac{f_n^2}{\beta_n} \leq \| f \|_{K_{B,\Phi}}^2 \sup_{n \geq 1} \frac{\beta_n}{\alpha_n} < \infty.
\]
(3.5)
Consequently, \( H(K_{B,\Phi}) \subset H(K_{A,\Phi}) \). Suppose then that \( H(K_{B,\Phi}) \subset H(K_{A,\Phi}) \) and assume to the contrary that \( \sup_{n \geq 1} \alpha_n^{-1} \beta_n = \infty \) so that there is a subsequence \( (n_m)_{m=1}^{\infty} \) such that \( \alpha_n^{-1} \beta_n \geq 2^m \).
Then \( f = \sum_{m=1}^{\infty} 2^{-m/2} \sqrt{\beta_{n_m}} \phi_{n_m} \in H(K_{B,\Phi}) \) since
\[
\| f \|_{K_{B,\Phi}}^2 = \sum_{m=1}^{\infty} \left( \frac{\sqrt{\beta_{n_m}}}{2^{m/2}} \right)^2 \frac{1}{\beta_{n_m}} = \sum_{m=1}^{\infty} 2^{-m} = 1
\]
but
\[
\| f \|_{K_{A,\Phi}}^2 = \sum_{m=1}^{\infty} \frac{2^{-m} \beta_{n_m}}{\alpha_{n_m}} \geq \sum_{m=1}^{\infty} 1 = \infty,
\]
which contradicts the assumption that \( H(K_{B,\Phi}) \subset H(K_{A,\Phi}) \). Therefore \( \sup_{n \geq 1} \alpha_n^{-1} \beta_n < \infty \). The second statement follows by setting \( \beta_n = 1 \) for every \( n \in \mathbb{N} \) and noting that then \( H(K_{B,\Phi}) = H(K) \).

Two normed spaces \( F \) and \( G \) are said to be norm-equivalent if they are equal as sets and if there exist positive constants \( C_1 \) and \( C_2 \) such that
\[
C_1 \| f \|_F \leq \| f \|_G \leq C_2 \| f \|_F \quad \text{for every} \quad f \in F.
\]
From (3.5) it can be concluded that \( H(K_{A,\Phi}) \) and \( H(K_{B,\Phi}) \) are norm-equivalent if \( \alpha_n \asymp \beta_n \).

Corollary 3.4. Let \( \Phi = (\phi_n)_{n=1}^{\infty} \) be an orthonormal basis of \( H(K) \) and \( A = (\alpha_n)_{n=1}^{\infty} \) and \( B = (\beta_n)_{n=1}^{\infty} \) two \( \Phi \)-scalings of \( H(K) \). Then \( H(K_{A,\Phi}) \) and \( H(K_{B,\Phi}) \) are norm-equivalent if and only if \( \alpha_n \asymp \beta_n \).

Proposition 3.5. Let \( \Phi = (\phi_n)_{n=1}^{\infty} \) be an orthonormal basis of \( H(K) \) and \( A = (\alpha_n)_{n=1}^{\infty} \) and \( B = (\beta_n)_{n=1}^{\infty} \) two \( \Phi \)-scalings of \( H(K) \). Then \( H(K_{B,\Phi}) \subseteq H(K_{A,\Phi}) \) if and only if \( \sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty \) and \( \beta_n \leq \alpha_n \).

Proof. Assume first that \( \sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty \) and \( \beta_n \leq \alpha_n \). Since \( \beta_n \leq \alpha_n \), Proposition 3.3 yields \( H(K_{B,\Phi}) \subset H(K_{A,\Phi}) \). Therefore \( H(K_{B,\Phi}) \) is a proper subset of \( H(K_{A,\Phi}) \) if \( H(K_{A,\Phi}) \) is not a subset of \( H(K_{B,\Phi}) \). But, again by Proposition 3.3, \( H(K_{A,\Phi}) \subset H(K_{B,\Phi}) \) if and only if \( \alpha_n \leq \beta_n \),
which contradicts the assumption that \(\sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty\). Hence \(H(K_{B, \Phi})\) is a proper subset of \(H(K_{A, \Phi})\).

Assume then that \(H(K_{B, \Phi}) \subseteq H(K_{A, \Phi})\). It follows from Proposition 3.3 that \(\beta_n \leq \alpha_n\). If \(\sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty\) did not hold, there would exist \(C > 0\) such that \(\alpha_n \leq C \beta_n\) for all \(n \in \mathbb{N}\), which is to say \(\alpha_n \leq \beta_n\). But by Proposition 3.3 this would imply that \(H(K_{B, \Phi}) \subset H(K_{A, \Phi})\), which would contradict the assumption that \(H(K_{B, \Phi})\) is a proper subset of \(H(K_{A, \Phi})\). This completes the proof. \(\square\)

**Remark 3.6.** Let \(H(R)\) be another separable RKHS of functions on \(T\). The RKHSs \(H(K)\) and \(H(R)\) are simultaneously diagonalisable if there exists an orthonormal basis \((\phi_n)_{n=1}^{\infty}\) of \(H(K)\) which is an orthogonal basis of \(H(R)\). That is, \((\|\phi_n\|^{-2})_{n=1}^{\infty}\) is an orthonormal basis of \(H(R)\) and consequently \(H(R) = H(K_{A, \Phi})\) for the scaling with \(\alpha_n = \|\phi_n\|^{-2}\).

We conclude this section by demonstrating that scaled RKHSs generalise powers of RKHSs. We say that a \(\Phi\)-scaling \(A_{\rho} = (\alpha_n)_{n=1}^{\infty}\) of \(H(K)\) is \(\rho\)-hyperharmonic if \(\alpha_n = n^\rho\) for some \(\rho \geq 0\). The RKHS \(H(K_{A_{\rho}, \Phi})\) is a \(\rho\)-hyperharmonic scaled RKHS. Recall from Section 2 that if \(T\) is a compact metric space, \(K\) is continuous on \(T \times T\), and \(\nu\) is a finite and strictly positive Borel measure on \(T\), then the integral operator (2.1) has eigenfunctions \((\psi_n)_{n=1}^{\infty}\) and decreasing positive eigenvalues \((\lambda_n)_{n=1}^{\infty}\) and \(\Psi = (\sqrt{\lambda_n} \psi_n)_{n=1}^{\infty}\) is an orthonormal basis of \(H(K)\). For \(\theta > 0\) the kernel \(K^{(\theta)}(t, t') = \sum_{n=1}^{\infty} \lambda_n^\theta \psi_n(t) \psi_n(t')\) is the \(\theta\)th power of \(K\) and its RKHS \(H(K^{(\theta)})\) the \(\theta\)th power of \(H(K)\). These objects are well-defined if \(\sum_{n=1}^{\infty} \lambda_n^\theta \psi_n(t) < \infty\) for all \(t \in T\). We immediately recognise that \(K^{(\theta)}\) equals the scaled kernel \(K_{A_{\theta}, \Psi}\) for the scaling \(A = (\lambda_n^\theta)_{n=1}^{\infty}\) because

\[
K_{A_{\theta}, \Psi}(t, t') = \sum_{n=1}^{\infty} \lambda_n^\theta \psi_n(t) \psi_n(t') = \sum_{n=1}^{\infty} \lambda_n^\theta \psi_n(t) \psi_n(t') = K^{(\theta)}(t, t').
\]

An important special case is when the eigenvalues have polynomial decay, \(\lambda_n \asymp n^{-p}\) for some \(p > 0\). For example, this holds, with \(p = -2s/d\), if \(T \subset \mathbb{R}^d\) and \(H(K)\) is norm-equivalent to the Sobolev space \(W_s^2(T)\) for \(s > d/2\) [36, p. 370]. If \(\lambda_n \asymp n^{-p}\) and \(\rho = p(1 - \theta)\), the \(\rho\)-hyperharmonic scaled RKHS \(H(K_{A_{\rho}, \Psi})\) is norm-equivalent to the \(\theta\)th power RKHS \(H(K^{(\theta)})\) by Corollary 3.4 because \(n^\rho \lambda_n \asymp n^{-\theta p} \asymp \lambda_n^\theta\).

### 4. Sample Path Properties

This section contains the main results of the article. We begin by considering a specialisation to scaled RKHSs of a theorem originally proved by Driscoll [9] and later generalised by Lukić and Beder [24]. Then we define general sample support sets and proceed to characterise them for \(\sigma\)-algebras generated by scalings of \(H(K)\).

#### 4.1. Domination and Generalised Driscoll’s Theorem for Scaled RKHSs

A kernel \(R\) on \(T\) dominates \(K\) if \(H(K) \subset H(R)\). In this case there exists [24, Theorem 1.1] a unique linear operator \(L : H(R) \to H(K)\), called the dominance operator, whose range is contained in \(H(K)\) and which satisfies

\[
\langle f, g \rangle_R = \langle Lf, g \rangle_K \quad \text{for all} \quad f \in H(R) \quad \text{and} \quad g \in H(K).
\]
The dominance is said to be nuclear, denoted $R \gg K$, if $H(R)$ is separable and the operator $L$ is nuclear, which is to say that

$$\text{tr}(L) = \sum_{n=1}^{\infty} (L\psi_n, \psi_n)_R < \infty$$

(4.1)

for any orthonormal basis $(\psi_n)_{n=1}^{\infty}$ of $H(R)$.\(^3\)

Define the pseudometric

$$d_R(t, t') = \|R(\cdot, t) - R(\cdot, t')\|_R = \sqrt{R(t, t) - 2R(t, t') + R(t', t')}$$

on $T$. If $R$ is positive-definite, $d_R$ is a metric. However, positive-definiteness is not necessary for $d_R$ to be a metric. For example, the Brownian motion kernel $R(t, t') = \min\{t, t'\}$ on $T = [0, 1]$ is positive-semidefinite because $R(t, 0) = 0$ for every $t \in T$ but nevertheless yields a metric since $d_R(t, t') = \sqrt{t-2\min\{t, t'\} + t'}$ vanishes if and only if $t = t'$. See [24, Section 4] for more properties of $d_R$. An equivalent condition to $d_R$ being a metric is to require that the mapping $t \mapsto R(\cdot, t)$ be injective.

According to the following theorem, a special case of the zero-one law of Kallianpur [15, 21] and a generalisation by Lukić and Beder [24, Theorem 7.5] of an earlier result by Driscoll [9, Theorem 3], the nuclear dominance condition determines whether or not samples of a Gaussian process $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ lie in $H(R)$. In particular, the probability of them being in $H(R)$ is always either one or zero.

**Theorem 4.1 (Generalised Driscoll’s Theorem).** Let $R$ be a continuous kernel on $T \times T$ with separable RKHS and $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$. If $d_R$ is a metric, then either

$$\mathbb{P}[X \in H(R)] = 0 \quad \text{and} \quad R \not\gg K$$

or

$$\mathbb{P}[X \in H(R)] = 1 \quad \text{and} \quad R \gg K.$$

**Proof.** Theorem 7.5 in [24] is otherwise identical except that $R$ is not assumed $d_T$-continuous and the samples of $X$ are assumed $d_R$ continuous. However, when $R$ is $d_T$-continuous, $d_T$-continuity implies $d_R$-continuity. \(\square\)

Summability of the reciprocal scaling coefficients controls whether or not a scaled RKHS contains the sample paths.

**Lemma 4.2.** Let $\Phi = (\phi_n)_{n=1}^{\infty}$ be an orthonormal basis of $H(K)$ and $A = (\alpha_n)_{n=1}^{\infty}$ a $\Phi$-scaling of $H(K)$. Let $R = K_{A, \Phi}$. If $d_K$ is a metric, then so is $d_R$.

**Proof.** Because $d_K$ is a metric,

$$d_K(t, t')^2 = K(t, t) - 2K(t, t') + K(t', t') = \sum_{n=1}^{\infty} |\phi_n(t) - \phi_n(t')|^2$$

vanishes if and only if $t = t'$. Since $d_R(t, t')^2 = \sum_{n=1}^{\infty} \alpha_n |\phi_n(t) - \phi_n(t')|^2$ and $\alpha_n$ are positive, we conclude that $d_R(t, t') = 0$ if and only if $t = t'$. \(\square\)

\(^3\)A change of basis shows that $\text{tr}(L)$ does not depend on the orthonormal basis.
Theorem 4.3. Let $\Phi = (\phi_n)_{n=1}^{\infty}$ be an orthonormal basis of $H(K)$, $A = (\alpha_n)_{n=1}^{\infty}$ a $\Phi$-scaling of $H(K)$, and $(X(t))_{t \in T} \sim GP(0, K)$. If $K_{A, \Phi}$ is continuous and $d_K$ is a metric, then either

$$\mathbb{P}[X \in H(R)] = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{\alpha_n} = \infty$$

or

$$\mathbb{P}[X \in H(R)] = 1 \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{\alpha_n} < \infty.$$ 

Proof. Assume first that the scaling is such that $H(K) \subset H(K_{A, \Phi})$. It is easy to verify using Proposition 3.2 that the unique dominance operator $L : H(K_{A, \Phi}) \to H(K)$ is given by

$$Lf = \sum_{n=1}^{\infty} \frac{f_n}{\alpha_n} \phi_n \quad \text{for any} \quad f = \sum_{n=1}^{\infty} f_n \phi_n \in H(K_{A, \Phi}).$$

Since $(\sqrt{\alpha_n} \phi_n)_{n=1}^{\infty}$ is an orthonormal basis of $H(K_{A, \Phi})$ and $L(\sqrt{\alpha_n} \phi_n) = \alpha_n^{-1/2}$, the nuclear dominance condition (4.1) is

$$\text{tr}(L) = \sum_{n=1}^{\infty} \langle \alpha_n L \phi_n, \sqrt{\alpha_n} \phi_n \rangle_{K_{A, \Phi}} = \sum_{n=1}^{\infty} \langle \phi_n, \phi_n \rangle_{K_{A, \Phi}} = \sum_{n=1}^{\infty} \frac{1}{\alpha_n},$$

and the claim thus follows from Theorem 4.1 since Lemma 4.2 guarantees that $d_R$ is a metric when $R$ is obtained by scaling $K$. Then, if $H(K) \not\subset H(K_{A, \Phi})$ it is trivially true that $K_{A, \Phi} \not\gg K$ and to complete the proof we have to show that this implies $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$. If we had $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$, then it would necessarily hold that $\sup_{n \geq 1} \alpha_n^{-1} < \infty$ and consequently $H(K) \subset H(K_{A, \Phi})$ by Proposition 3.3. Therefore $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$. \qed

4.2. Sample Support Sets

Theorems 4.1 and 4.3 motivate us to define the sample support set of a Gaussian process with respect to a collection of kernels as the largest set on the “boundary” between their induced RKHSs of probabilities one and zero. Let $\mathcal{R}$ be a collection of continuous kernels $R$ on $T$ for which $d_R$ is a metric and $H(\mathcal{R})$ the corresponding set of RKHSs. Every element of $H(\mathcal{R})$ is a subset of $C(T)$. By the generalised Driscoll’s theorem each element of $H(\mathcal{R})$ has $\mu_X$-measure one or zero, depending on the nuclear dominance condition. Define the disjoint sets

$$\mathcal{R}_1(K) = \{ R \in \mathcal{R} : R \gg K \} \quad \text{and} \quad \mathcal{R}_0(K) = \{ R \in \mathcal{R} : R \not\gg K \}$$

which partition $\mathcal{R}$. We assume that both $\mathcal{R}_1(K)$ and $\mathcal{R}_0(K)$ are non-empty and introduce the notion of a sample support set.

Definition 4.4 (Sample support set). Let $S(\mathcal{R}) = \sigma(H(\mathcal{R}))$ be the $\sigma$-algebra generated by $H(\mathcal{R})$. The sample support set, $S_{\mathcal{R}}(K)$, of the Gaussian process $(X(t))_{t \in T} \sim GP(0, K)$ with respect to $\mathcal{R}$ is the largest subset of $C(T)$ such that

$$S_{\mathcal{R}}(K) \subset H \quad \text{for every} \quad H \in S(\mathcal{R}) \quad \text{such that} \quad \mu_X(H) = 1.$$
Proposition 4.5.  It holds that

$$S_{\mathfrak{R}}(K) = \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \setminus \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0).$$

(4.2)

Proof. Suppose that there is \( f \in S_{\mathfrak{R}}(K) \) which is not contained in the set on the right-hand side of (4.2). That is, we have either that

$$f \notin \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \quad \text{or} \quad f \in \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0).$$

In the former case there is \( R_1 \in \mathfrak{R}_1(K) \) such that \( f \notin H(R_1) \). But because \( \mu_X(H(R_1)) = 1 \) and \( S_{\mathfrak{R}}(K) \subset H(R_1) \) by definition, this violates the assumption that \( f \in S_{\mathfrak{R}}(K) \). In the latter case there is \( R_0 \in \mathfrak{R}_0(K) \) such that \( f \in H(R_0) \). As \( \mu_X(H(R_0)) = 0 \), we have for any \( R_1 \in \mathfrak{R}_1 \) that \( \mu_X(H(R_1) \setminus H(R_0)) = 1 \). But since \( f \notin H(R_1) \setminus H(R_0) \), the assumption that \( f \in S_{\mathfrak{R}}(K) \) is again violated and we conclude that \( S_{\mathfrak{R}}(K) \subset \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \setminus \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0) \).

Since all elements of \( H(\mathfrak{R}) \) are either of measure zero or one, so are those of \( S(\mathfrak{R}) \). It is therefore clear that \( \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \setminus \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0) \) is contained in every \( H \in S(\mathfrak{R}) \) such that \( \mu_X(H) = 1 \). Consequently, \( \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \setminus \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0) \subset S_{\mathfrak{R}}(K) \). This concludes the proof. \( \square \)

The sample support set is the largest set which is contained in every set of probability one under the law of \( X \) that can be expressed in terms of countably many elementary set operations of the RKHSs \( H(R) \) for \( R \in \mathfrak{R} \). The larger \( \mathfrak{R} \) is, the more precisely \( S_{\mathfrak{R}}(K) \) describes the samples of \( X \). But there is an important caveat. If \( \mathfrak{R} \) is countable, the sample support set is in the \( \sigma \)-algebra \( B \), defined in (1.3), and has \( \mu_X \)-measure one. However, when \( \mathfrak{R} \) is uncountable and does not contain countable subsets \( \mathfrak{R}_1(K) \subset \mathfrak{R}_1(K) \) and \( \mathfrak{R}_0(K) \subset \mathfrak{R}_0(K) \) such that

$$\bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) = \bigcap_{R_1 \in \mathfrak{R}'_1(K)} H(R_1) \quad \text{and} \quad \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0) = \bigcup_{R_0 \in \mathfrak{R}'_0(K)} H(R_0),$$

it cannot be easily determined if \( S_{\mathfrak{R}}(K) \) is an element of \( B \).

In this article we are mainly interested in sample support sets with respect to \( \mathfrak{R} \) which consists of all scaled kernels (and will in Theorem 4.8 characterise this set). It is nevertheless conceivable that one may want to or be forced to work with less rich set of kernels—scaled or not—and with such eventuality in mind we have introduced the more general concept of a sample support set. If \( \mathfrak{R} \) is a collection of scaled kernels, the sample support set takes a substantially more concrete form. For this purpose we introduce the concept an approximately constant sequence, which is inspired by the results collected in [19, § 41].

Definition 4.6 (Approximately constant sequence).  Let \( \Sigma \) be a collection of non-negative sequences. A non-negative sequence \( (a_n)_{n=1}^{\infty} \) is said to be \( \Sigma \)-approximately constant if for every \( (b_n)_{n=1}^{\infty} \in \Sigma \) the series \( \sum_{n=1}^{\infty} b_n \) and \( \sum_{n=1}^{\infty} a_n b_n \) either both converge or diverge.

We mention two basic properties of approximately constant sequences: (i) If \( (a_n)_{n=1}^{\infty} \) and \( (a'_n)_{n=1}^{\infty} \) are two \( \Sigma \)-approximately constant sequences, then so is their sum \( (a_n + a'_n)_{n=1}^{\infty} \). (ii) The larger \( \Sigma \) is, the fewer \( \Sigma \)-approximately constant sequences there are. That is, if \( \Sigma_1 \) and \( \Sigma_2 \) are two collections of non-negative sequences such that \( \Sigma_1 \subset \Sigma_2 \), then a non-negative sequence \( a \) is \( \Sigma_1 \)-approximately constant if it is \( \Sigma_2 \)-approximately constant.

For the RKHS \( H(K) \) and any of its orthonormal basis \( \Phi = (\phi_n)_{n=1}^{\infty} \) we let \( \mathcal{R}(\Sigma, \Phi) \) denote the set of all functions \( f = \sum_{n=1}^{\infty} f_n \phi_n \) such that the series converges pointwise on \( T \) and \( (f_n^2)_{n=1}^{\infty} \) is a
\(\Sigma\)-approximately constant sequence. The following theorem provides a crucial connection between sample support sets with respect to scaled kernels and functions defined as orthonormal expansions with approximately constant coefficients.

**Theorem 4.7.** Let \(\Phi = (\phi_n)_{n=1}^{\infty}\) be an orthonormal basis of \(H(K)\) and \(\Sigma_{\Phi}\) a collection of \(\Phi\)-scalings of \(H(K)\) such that the corresponding scaled kernels are continuous. Suppose that \(d_K\) is a metric and let \(\mathcal{R} = \{ K_{A,\phi} : A \in \Sigma_{\Phi} \}. \) Then \(S_{\mathcal{R}}(K) = \mathcal{R}(\Sigma_{\Phi}, \Phi)\).

**Proof.** Note first that, by Lemma 4.2, \(d_R\) is a metric for every \(R \in \mathcal{R}\). Because every scaling of \(H(K)\) has an orthonormal basis that is a scaled version of \((\phi_n)_{n=1}^{\infty}\), every \(f \in S_{\mathcal{R}}(K)\) can be written as \(f = \sum_{n=1}^{\infty} f_n \phi_n\) for some real coefficients \(f_n\). Let \(\Sigma_1(K)\) and \(\Sigma_0(K)\) stand for the collections of \((\alpha_n)_{n=1}^{\infty} \in \Sigma_{\Phi}\) such that \(\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty\) and \(\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty\), respectively. Then, by Theorem 4.3, \(K_{A,\phi} \in \mathcal{R}_1(K)\) if \(A \in \Sigma_1(K)\) and \(K_{A,\phi} \in \mathcal{R}_0(K)\) if \(A \in \Sigma_0(K)\). Because, by definition, \(S_{\mathcal{R}}(K) \subset H(K_{A,\phi})\) for any \(A \in \Sigma_1(K)\) and \(S_{\mathcal{R}}(K) \cap H(K_{A,\phi}) = \emptyset\) for any \(A \in \Sigma_0(K)\) it follows that for every \(f \in S_{\mathcal{R}}(K)\) and any \((\alpha_n)_{n=1}^{\infty} \in \Sigma_{\Phi}\) we have

\[
\sum_{n=1}^{\infty} \frac{f_n^2}{\alpha_n} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{\alpha_n} < \infty \quad \text{or} \quad \sum_{n=1}^{\infty} \frac{f_n^2}{\alpha_n} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{\alpha_n} = \infty.
\]

That is, \((f_n^2 \alpha_n)_{n=1}^{\infty}\) is a \(\Sigma_{\Phi}\)-approximately constant sequence and thus \(S_{\mathcal{R}}(K) \subset \mathcal{R}(\Sigma_{\Phi}, \Phi)\). Conversely, if \(f \in \mathcal{R}(\Sigma_{\Phi}, \Phi)\), then \(f \in H(K_{A,\phi})\) for every \(A \in \Sigma_1(K)\) and \(f \notin H(K_{A,\phi})\) for every \(A \in \Sigma_0(K)\). Hence \(f \in S_{\mathcal{R}}(K)\) and thus \(S_{\mathcal{R}}(K) = \mathcal{R}(\Sigma_{\Phi}, \Phi)\). \(\square\)

Next we use Theorem 4.7 to describe the sample support set more concretely.

### 4.3. Sample Support Sets for Scaled RKHSs

Let \(\Sigma\) be the set of all positive sequences. Then the collection of \(\Sigma\)-approximately constant sequences is precisely the collection of non-negative sequences \((a_n)_{n=1}^{\infty}\) such that

\[
\liminf_{n \to \infty} a_n > 0 \quad \text{and} \quad \sup_{n \geq 1} a_n < \infty. \quad (4.3)
\]

For suppose that there were a \(\Sigma\)-approximately constant sequence \((a_n)_{n=1}^{\infty}\) that violated (4.3). If \(\liminf_{n \to \infty} a_n = 0\), then there is a subsequence \((a_{nm})_{m=1}^{\infty}\) such that \(a_{nm} \leq 2^{-m}\) for all \(m \in \mathbb{N}\). Let \((b_n)_{n=1}^{\infty} \in \Sigma\) be a sequence such that \(b_n = 2^{-n} a_n^{-1}\) for \(n \notin (n_m)_{m=1}^{\infty}\) and \(b_{nm} = 1\) for \(m \in \mathbb{N}\). Then \(\sum_{n=m}^{\infty} b_n\) diverges but

\[
\sum_{n=1}^{\infty} a_n b_n = \sum_{n \notin (n_m)_{m=1}^{\infty}} a_n b_n + \sum_{m=1}^{\infty} a_{nm} b_{nm} \leq \sum_{n \notin (n_m)_{m=1}^{\infty}} 2^{-n} + \sum_{m=1}^{\infty} 2^{-m} < \infty,
\]

which contradicts the assumption that \((a_n)_{n=1}^{\infty}\) is a \(\Sigma\)-approximately constant sequence. An analogous argument with \(a_{nm} \geq 2^m\), \(b_n = 2^{-n}\), and \(b_{nm} = 2^{-m}\) shows that the second condition in (4.3) cannot be violated either and thus every \(\Sigma\)-approximately constant sequence satisfies (4.3). It is trivial that a sequence satisfying (4.3) is \(\Sigma\)-approximately constant because the conditions imply the existence of positive constants \(c_1 \leq c_2\) such that \(c_1 \leq a_n \leq c_2\) for all sufficiently large \(n\). This, together with Theorem 4.7, yields the following theorem which we consider the main result of this article. The proof of this theorem is somewhat more complicated than the above argument because we cannot assume that every positive sequence is a scaling of \(H(K)\).
Theorem 4.8. Let $\Phi = (\phi_n)_{n=1}^{\infty}$ be an orthonormal basis of $H(K)$ and suppose that there is a $\Phi$-scaling $A = (\alpha_n)_{n=1}^{\infty}$ of $H(K)$ such that $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$ and $K_{A,\Phi}$ is continuous. Let $\Sigma_{\Phi}$ be the collection $\Phi$-scalings of $H(K)$ such that the corresponding scaled kernels are continuous. Suppose that $d_K$ is a metric and let $\mathcal{R} = \{K_{A,\Phi} : A \in \Sigma_{\Phi}\}$. Then $S_{\mathcal{R}}(K)$ is non-empty and consists precisely of the functions $f = \sum_{n=1}^{\infty} f_n \phi_n$ such that

$$\liminf_{n \to \infty} f_n^2 > 0 \quad \text{and} \quad \sup_{n \geq 1} f_n^2 < \infty, \quad (4.4)$$

Proof. By Theorem 4.7, we know that $S_{\mathcal{R}}(K) = \mathcal{R}(\Sigma_{\Phi}, \Phi)$ and the assumption that there exists a scaling $A = (\alpha_n)_{n=1}^{\infty}$ of $H(K)$ such that $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$ implies that $\mathcal{R}(\Sigma_{\Phi}, \Phi)$ is non-empty. Hence we have to show that functions in $\mathcal{R}(\Sigma_{\Phi}, \Phi)$ satisfy (4.4). Suppose that there is a function $f = \sum_{n=1}^{\infty} f_n \phi_n \in \mathcal{R}(\Sigma_{\Phi}, \Phi)$ that violates (4.4). If $\sup_{n \geq 1} f_n^2 = \infty$, then there is a subsequence $(f_{n_m})_{m=1}^{\infty}$ such that $f_{n_m}^2 \geq 2^m$ for all $m \in \mathbb{N}$. Define a sequence $B = (\beta_n)_{n=1}^{\infty}$ by setting $\beta_{n_m} = 2^m f_{n_m}^2$ for $m$ such that $f_{n_m}^2 < \alpha_{n_m}$ and $\beta_n = \alpha_n$ for all other $n$. Then $B$ is a $\Phi$-scaling of $H(K)$ since $\beta_n \leq \alpha_n$ and $\sum_{n=1}^{\infty} \beta_n^{-1} \leq \sum_{n=1}^{\infty} \alpha_n^{-1} + \sum_{m=1}^{\infty} 2^{-m} < \infty$. Moreover, Proposition 3.3 implies that $H(K_{B,\Phi}) \subset H(K_{A,\Phi})$ and it thus follows from the continuity of $K_{A,\Phi}$ that $K_{B,\Phi}$ is continuous, so that $B \in \Sigma_{\Phi}$. However,

$$\sum_{n=1}^{\infty} \frac{f_n^2}{\beta_n} = \sum_{n \notin \{n_m\}_{m=1}^{\infty}} \frac{f_n^2}{\beta_n} + \sum_{m \in \mathbb{N}} \frac{f_{n_m}^2}{\beta_{n_m}} + \sum_{m \in \mathbb{N}} \frac{f_{n_m}^2}{\beta_{n_m}} \geq \sum_{n \notin \{n_m\}_{m=1}^{\infty}} \frac{f_n^2}{\alpha_n} + \sum_{m \in \mathbb{N}} \frac{f_{n_m}^2}{\alpha_{n_m}} + \sum_{m \in \mathbb{N}} \frac{f_{n_m}^2}{\alpha_{n_m}} = \infty,$$

which contradicts the assumption that $(f_n^2)_{n=1}^{\infty}$ is $\Sigma_{\Phi}$-approximately constant. On the other hand, if $\liminf_{n \to \infty} f_n^2 = 0$, then there is a subsequence $(f_{n_m})_{m=1}^{\infty}$ such that $f_{n_m}^2 \leq 2^{-m}$ for all $m \in \mathbb{N}$. The sequence $B = (\beta_n)_{n=1}^{\infty}$ defined as $\beta_{n_m} = 1$ for $m \in \mathbb{N}$ and $\beta_n = \alpha_n f_n^2$ for other $n$ is a $\Phi$-scaling of $H(K)$ and $K_{B,\Phi} \in \Sigma_{\Phi}$ because we have proved that $\sup_{n \geq 1} f_n^2 < \infty$. Furthermore, clearly $\sum_{n=1}^{\infty} \beta_n^{-1} = \infty$ but

$$\sum_{n=1}^{\infty} \frac{f_n^2}{\beta_n} = \sum_{n \notin \{n_m\}_{m=1}^{\infty}} \frac{f_n^2}{\beta_n} + \sum_{m=1}^{\infty} \frac{f_{n_m}^2}{\beta_{n_m}} \leq \sum_{n \notin \{n_m\}_{m=1}^{\infty}} \frac{1}{\alpha_n} + \sum_{m=1}^{\infty} 2^{-m} < \infty,$$

which again contradicts the assumption that $(f_n^2)_{n=1}^{\infty}$ is $\Sigma_{\Phi}$-approximately constant. This completes the proof. \square
Then yields, by Lemma 4.9, a convergent series such that \( \sum_{n=1}^{\infty} f_n \varphi_n = \sum_{n=1}^{\infty} f_{\pi(n)} \varphi_{\pi(n)} \) but it is clear that

\[
\sup_{n \geq 1} f_n^2 = \sup_{n \geq 1} f_{\pi(n)}^2 \quad \text{and} \quad \liminf_{n \to \infty} f_n^2 = \liminf_{n \to \infty} f_{\pi(n)}^2,
\]

so that which of the bases \( (\varphi_n)_{n=1}^{\infty} \) and \( (\varphi_{\pi(n)})_{n=1}^{\infty} \) the function \( f \) is expanded in does not matter.

Theorem 4.11 below demonstrates that the sample support set of Theorem 4.8 cannot be expressed in terms of countably many elementary set operations of scaled RKHSs—and thus could be non-measurable. The reason for this is that there does not exist a useful notion of a “boundary” between convergent and divergent series, a result which is partially contained in Lemma 4.10. The lemma is a modified version of a result originally due to du Bois-Reymund and Hadamard [19, pp. 301–302]. In its proof a classical result by Dini [19, p. 293] is needed.

**Lemma 4.9 (Dini).** Let \( (a'_n)_{n=1}^{\infty} \) be a positive sequence such that \( \sum_{n=1}^{\infty} a'_n < \infty \) and set

\[
a_n = \frac{a'_n}{\left(\sum_{l=n}^{\infty} a'_l\right)^{1/2}}.
\]

Then \( \lim_{n \to \infty} (a'_n)^{-1} a_n = \infty \) and the series \( \sum_{n=1}^{\infty} a_n \) converges if and only if \( 0 \leq c < 1 \).

**Lemma 4.10.** For each \( m \in \mathbb{N} \), suppose that \( (a_{m,n})_{n=1}^{\infty} \) is a positive sequence such that \( \sum_{n=1}^{\infty} a_{m,n} < \infty \). Then there is a positive sequence \( (a_n)_{n=1}^{\infty} \) such that

\[
\sum_{n=1}^{\infty} a_n < \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{a_n}{a_{m,n}} = \infty \quad \text{for every } m \in \mathbb{N}.
\]

**Proof.** For each \( m \in \mathbb{N} \), define the sequence \( (\bar{a}_{m,n})_{n=1}^{\infty} \) by setting \( \bar{a}_{m,n} = \sum_{k=1}^{m} a_{k,n} \), so that \( \bar{a}_{m,n} \leq \bar{a}_{m+1,n} \) and \( \sum_{n=1}^{\infty} \bar{a}_{m,n} < \infty \) for every \( m \). Therefore there are \( n_m \) such that \( n_m < n_{m+1} \) and \( \sum_{n=n_m}^{\infty} \bar{a}_{m,n} \leq 2^{-m} \) for every \( m \geq 2 \). Define \( a'_n = \bar{a}_{1,n} \) when \( n < n_2 \) and \( a'_n = a_{m,n} \) when \( n_m \leq n < n_{m+1} \) for \( m \geq 2 \). Then

\[
\sum_{n=1}^{\infty} a'_n = \sum_{n=1}^{n_2-1} \bar{a}_{1,n} + \sum_{m=2}^{\infty} \sum_{n=n_m}^{n_{m+1}-1} \bar{a}_{m,n} \leq \sum_{n=1}^{\infty} \bar{a}_{1,n} + \sum_{m=2}^{\infty} 2^{-m} < \infty
\]

and, because \( \bar{a}_{m,n} \leq \bar{a}_{m+1,n} \) and \( \bar{a}_{m,n} \geq a_{m,n} \), we have \( a'_n \geq \bar{a}_{m,n} \geq a_{m,n} \) for all \( m \) and \( n \geq n_m \).

Finally, selecting

\[
a_n = \frac{a'_n}{\left(\sum_{l=n}^{\infty} a'_l\right)^{1/2}}
\]

yields, by Lemma 4.9, a convergent series such that

\[
\lim_{n \to \infty} \frac{a_n}{a_{m,n}} \leq \lim_{n \to \infty} \frac{a_n}{a_{n}} = \lim_{n \to \infty} \left(\sum_{l=n}^{\infty} a'_l\right)^{-1/2} = \infty.
\]

\( \Box \)
**Theorem 4.11.** Let $\Phi = (\phi_n)_{n=1}^\infty$ be an orthonormal basis of $H(K)$ and suppose that there is a $\Phi$-scaling $A = (\alpha_n)_{n=1}^\infty$ of $H(K)$ such that $\sum_{n=1}^\infty \alpha_n^{-1} < \infty$ and $K_{A,\Phi}$ is continuous. Let $\Sigma_\Phi$ be the collection $\Phi$-scalings of $H(K)$ such that the corresponding scaled kernels are continuous. Suppose that $d_K$ is a metric and let $\mathcal{R} = \{ K_{A,\Phi} : A \in \Sigma_\Phi \}$. Let $A_m = (\alpha_{m,n})_{n=1}^\infty$ and $B_m = (\beta_{m,n})_{n=1}^\infty$ for $m \in \mathbb{N}$ be any elements of $\Sigma_\Phi$ such that

$$\sum_{n=1}^\infty \frac{1}{\alpha_{m,n}} < \infty \quad \text{and} \quad \sum_{n=1}^\infty \frac{1}{\beta_{m,n}} = \infty \quad \text{for every } m \in \mathbb{N}.$$  

If $(X(t))_{t \in T} \sim GP(0, K)$, then there exists $F \in \mathcal{B}$ such that $\mu_X(F) = 1$ and

$$S_{\mathcal{R}}(K) \subseteq F \subseteq \bigcap_{m=1}^\infty H(K_{A_m, \Phi}) \setminus \bigcup_{m=1}^\infty H(K_{B_m, \Phi}).$$

**Proof.** By Lemma 4.10 there is a positive sequence $A' = (\alpha'_{n})_{n=1}^\infty$ such that

$$\sum_{n=1}^\infty \frac{1}{\alpha'_{n}} < \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{\alpha_{m,n}}{\alpha'_{n}} = \infty \quad \text{for every } m \in \mathbb{N}.$$  

By Proposition 3.5, $H(K_{A', \Phi}) \subseteq H(K_{A_m, \Phi})$ for every $m$. Therefore $H(K_{A', \Phi}) \subset \bigcap_{m=1}^\infty H(K_{A_m, \Phi})$. Using Lemma 4.9 and Proposition 3.5 we can construct a $\Phi$-scaling $A$ such that $H(K_{A, \Phi}) \subseteq H(K_{A', \Phi})$ and hence the set $F = H(K_{A, \Phi}) \setminus \bigcup_{m=1}^\infty H(K_{B_m, \Phi})$ has the desired properties by Theorem 4.3.\[\Box\]

In Theorem 4.11 the set $F$ was constructed by finding a scaled RKHS which is smaller than each $H(K_{A_m, \Phi})$ while still containing the sample set. A natural question is if we can also construct a scaled RKHS that contains each $H(K_{B_m, \Phi})$ but not the sample support set. Although Lemma 4.10 has a counterpart for divergent series, it seems difficult to guarantee that the resulting sequence is a scaling. Namely, it can be shown that if $(\beta_{m,n})_{n=1}^\infty$ in Theorem 4.11 are such that

$$\lim_{n \to \infty} \frac{\beta_{m,n}}{\beta_{m+1,n}} = 0 \quad \text{for every } m \in \mathbb{N}$$

then there exists $B = (\beta_n)_{n=1}^\infty$ such that

$$\sum_{n=1}^\infty \frac{1}{\gamma_n} = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{\beta_{m,n}}{\beta_{n}} = 0 \quad \text{for every } m \in \mathbb{N}.$$  

However, even though we know that there is a $\Phi$-scaling of $H(K)$ such that $\sum_{n=1}^\infty \alpha_n^{-1} < \infty$, the fact that $\sum_{n=1}^\infty \beta_n^{-1}$ diverges does not necessarily imply that $B$ is a $\Phi$-scaling because a divergent series can have terms that are arbitrarily larger than those of a convergent series [19, p. 303]: there exists a positive sequence $(\gamma_n)_{n=1}^\infty$ such that

$$\sum_{n=1}^\infty \frac{1}{\gamma_n} < \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{\gamma_{m,n}}{\beta_{n}} = 0.$$  

If $(\gamma_n)_{n=1}^\infty$ is a $\Phi$-scaling of $H(K)$, we cannot directly deduce that so is $B$ because $\beta_n \preceq \gamma_n$ does not hold.
4.4. On Monotonely Scaled RKHSs

In this section we demonstrate that the characterisation (4.4) fails if only monotone scalings are used. We say that a scaling \( A = (a_n)_{n=1}^\infty \) is monotone if \((a_n)_{n=1}^\infty \) is a monotone sequence. Let \( \Sigma_m \) be the collection of monotone \( \Phi \)-scalings of \( H(K) \) and \( R_m \) the collection of corresponding scaled kernels. Then the sample support set \( S_{R_m}(K) \) is larger than the sample support set \( S_{R}(K) \) characterised by Theorem 4.8 because, as we show below, a \( \Sigma_m \)-approximately constant sequence can be unbounded, as long as it does not grow too rapidly, and contain arbitrarily long sequences of zeros.

Let \((g_m)_{m=1}^\infty \) be a strictly increasing sequence of positive integers and set \( a_{gm} = g_{m+1} - g_m \) for \( m \geq 0 \) and \( a_0 = 0 \) otherwise. Then

\[
\sum_{n=1}^\infty a_n b_n = \sum_{m=0}^\infty (g_{m+1} - g_m) b_{gm}.
\]

(4.5)

Schlömilch’s generalisation of the Cauchy condensation test [19, p. 121] implies that, under the assumption \( g_{m+1} - g_m \leq C(g_m - g_{m-1}) \) for some \( C > 0 \) and all \( m \geq 1 \), for any non-increasing positive sequence \((b_n)_{n=1}^\infty \) the series \( \sum_{n=1}^\infty b_n \) and \( \sum_{m=0}^\infty (g_{m+1} - g_m) b_{gm} \) either both converge or diverge. That \((a_n)_{n=1}^\infty \) is \( \Sigma_m \)-approximately constant follows then from (4.5). The condition \( g_{m+1} - g_m \leq C(g_m - g_{m-1}) \), or equivalently \( a_{gm} \leq C a_{gm-1} \), guarantees that \( a_{gm} \) does not grow too fast in relation to \( g_m \). The canonical example is obtained by setting \( g_m = 2^m \) so that \( a_{2m} = 2^m \) and \( C = 2 \) and \( \sum_{m=0}^\infty (g_{m+1} - g_m) b_{gm} = \sum_{m=0}^\infty g_m b_{2m} \) converges or diverges with \( \sum_{n=1}^\infty b_n \) by the standard Cauchy condensation test. Note that, because

\[
\frac{1}{g_m} \sum_{n=1}^{g_m} a_n = \frac{1}{g_m} \sum_{k=0}^{m} (g_{k+1} - g_k) = \frac{g_{m+1} - g_0}{g_m} \leq \frac{(1 + C) g_m}{g_m} = 1 + C
\]

and

\[
\frac{1}{g_m - 1} \sum_{n=1}^{g_m-1} a_n = \frac{1}{g_m - 1} \sum_{k=0}^{m-1} (g_{k+1} - g_k) = \frac{g_m - g_0}{g_m - 1},
\]

the elements of the sequence \((a_n)_{n=1}^\infty \) are, on average, constants:

\[
\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a_n = 1 \quad \text{and} \quad \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^{N} a_n \leq 1 + C.
\]

This can be interpreted as a weaker form of (4.4).

Although \( \Sigma_m \) contains unbounded sequences and sequences with zero lower limit, no sequence in this set can have infinity or zero as its limit. This can be shown by using results in [19, § 41] which, given a sequence such that \( \lim_{n \to \infty} a_n = \infty \), guarantee the existence of a monotone sequence \((c_n)_{n=1}^\infty \) such that \( \sum_{n=1}^\infty c_n < \infty \) but \( \sum_{n=1}^\infty a_n c_n = \infty \). Conversely, given a sequence such that \( \lim_{n \to \infty} a_n = 0 \) it is possible to construct a monotone sequence \((d_n)_{n=1}^\infty \) such that \( \sum_{n=1}^\infty d_n = \infty \) but \( \sum_{n=1}^\infty a_n d_n < \infty \).

4.5. On Hyperharmonic Scalings and Iterated Logarithms
We conclude this section with two general constructions before moving onto concrete examples. Recall from Section 3 that scalings of the form \( A_\rho = (n^\rho)_{n=1}^\infty \) for any \( \rho \geq 0 \) are called hyperharmonic scalings. Because \( \sum_{n=1}^\infty n^{-\rho} \) converges if and only if \( \rho > 1 \), it follows from Theorem 4.3 that
\[
\mathbb{P}[X \in H(K_{A_\rho}, \Phi)] = 1 \text{ if } \rho > 1 \quad \text{and} \quad \mathbb{P}[X \in H(K_{A_\rho}, \Phi)] = 0 \text{ if } \rho \leq 1.
\]

Therefore the samples are “almost” contained in \( H(K_{A_\rho}, \Phi) \) while for any “small” \( \varepsilon > 0 \) the space \( H(K_{A_{1+\varepsilon}, \Phi}) \) is a “small” RKHS which contains the samples. Therefore the sample support set in Theorem 4.11 satisfies
\[
\text{Sup} = \bigcap_{K} \bigcup_{k=1}^{\infty} H(K_{A_{1+1/k}, \Phi}) \setminus H(K_{A_1}, \Phi),
\]
where the set on the right-hand side is in \( \mathcal{B} \) and has \( \mu_X \)-measure one. But Theorem 4.11 also guarantees the existence of a set \( F \in \mathcal{B} \) which is a proper subset of \( \bigcap_{k=1}^{\infty} H(K_{A_{1+1/k}, \Phi}) \setminus H(K_{A_1}, \Phi) \) while having \( \mu_X \)-measure one. One such set is \( F = H(K_{A_\rho}, \Phi) \setminus H(K_{A'}, \Phi) \) for scalings \( A = (\alpha_n)_{n=1}^\infty \) and \( A' = (\alpha'_n)_{n=1}^\infty \) defined by \( \alpha_n = n \log(n+1)^2 \) and \( \alpha'_n = n \log(n+1) \). This is because \( \sum_{n=1}^\infty \alpha_n^{-1} \) converges but \( \sum_{n=1}^\infty (\alpha'_n)^{-1} \) does not and
\[
\lim_{n \to \infty} \frac{n^\rho}{n \log(n+1)^2} = \infty \text{ for any } \rho > 1 \quad \text{and} \quad \lim_{n \to \infty} \frac{n \log(n+1)}{n} = \infty,
\]
which by Proposition 3.5 imply that \( H(K_{A_\rho}, \Phi) \subset \bigcap_{k=1}^{\infty} H(K_{A_{1+1/k}, \Phi}) \setminus H(K_{A_1}, \Phi) \) and \( H(K_{A_1}, \Phi) \subsetneq H(K_{A_\rho}, \Phi) \).

But one can construct even smaller measurable sets which contain the samples by the use of iterated logarithms. The iterated logarithm \( \log_p x \) is defined recursively as \( \log_p x = \log(\log_{p-1} x) \) for \( p \in \mathbb{N} \) and \( \log_0 x = x \). Let \( p \in \mathbb{N}_0 \) and suppose that \( q \geq 0 \) is large enough that \( \log_p(1+q) \) is positive. For any \( \rho > 1 \), define \( A = (\alpha_n)_{n=1}^\infty \) by
\[
\alpha_n = (n+q) \log(n+q) \times \cdots \times \log_{p-1}(n+q) \log_p(n+q)^\rho \quad (4.6)
\]
and \( A' = (\alpha'_n)_{n=1}^\infty \) by
\[
\alpha'_n = (n+q) \log(n+q) \times \cdots \times \log_{p-1}(n+q) \log_p(n+q). \quad (4.7)
\]
Then it can be proved that \( \sum_{n=1}^\infty \alpha_n^{-1} \) converges but \( \sum_{n=1}^\infty (\alpha'_n)^{-1} \) does not [19, pp. 123, 280, 293]. It is also clear that
\[
\lim_{n \to \infty} \frac{n \log(n+1)^2}{(n+q) \log(n+q) \times \cdots \times \log_{p-1}(n+q) \log_p(n+q)^\rho} = \infty
\]
and
\[
\lim_{n \to \infty} \frac{(n+q) \log(n+q) \times \cdots \times \log_{p-1}(n+q) \log_p(n+q)}{n \log(n+1)} = \infty
\]
if \( p \geq 2 \). Therefore iterated logarithms can be used to construct sets of \( \mu_X \)-measure one which are smaller than the set \( F \) above. But Theorem 4.11 again demonstrates that for any set of \( \mu_X \)-measure one constructed out of scalings defined by (4.6) and (4.7) there is a smaller set which still has measure one (and which contains the sample support set).
5. Examples

This section contains examples of kernels to which Theorem 4.3 can be applied to construct “small” RKHSs that contain the samples and “slightly smaller” RKHSs that do not. For simplicity we let the domain $T$ be a finite interval on the real line and occasionally index the orthonormal bases starting from zero. Note that to use our results one needs to have access to an orthonormal expansion of the kernel. This rules out examples involving the popular Matérn kernels because, to the best of our knowledge, no orthonormal expansions have been computed for these kernels. Previous results based on powers of RKHSs and convolution kernels are easier to apply in this regard, but are less flexible and expressive. See [36] and [17, Section 4.4] for examples featuring powers of RKHSs (in particular for Sobolev kernels and the Gaussian kernel) and [11, Appendices A.2 and A.3] for convolution kernel examples. Lukić [23] has examples involving integrated kernels.

5.1. Iterated Brownian Bridge Kernels

Let $T = [0, 1]$ and consider the iterated Brownian bridge kernel of integer order $s \geq 2$ [4, Section 4.1] defined as

$$K^s(t, t') = \frac{2}{\pi^s} \sum_{n=1}^{\infty} \frac{\sin(\pi n t) \sin(\pi n t')}{n^s} = \sum_{n=1}^{\infty} \phi_n(t) \phi_n(t'), \quad (5.1)$$

where $\phi_n(t) = \sqrt{2} (\pi n)^{-s} \sin(\pi n t)$. It can be shown that for even parameters the kernel (5.1) has the representation

$$K^{2s}(t, t') = (-1)^{s-1} \frac{2^{2s-1}}{(2s)!} \left[ B_{2s} \left( \frac{|t - t'|}{2} \right) - B_{2s} \left( \frac{t + t'}{2} \right) \right],$$

where $B_p$ is the Bernoulli polynomial of degree $p$. For $s = 2$ we obtain the Brownian bridge kernel

$$K^2(t, t') = \min\{t, t'\} - tt'.$$

For $s \geq 1$ and fixed $t' \in [0, 1]$ the even-order translates $K^{2s}(\cdot, t')$ are piecewise polynomials of order $2s - 1$.

Iterated Brownian bridge kernels are natural candidates for hyperharmonic scalings, as the $\rho$-hyperharmonic scaling $A_\rho = (n^\rho)^{\infty}_{n=1}$ for $\rho \in \mathbb{N}$ such that $\rho \leq s - 2$ gives

$$K^{s}_{A_\rho, \psi}(t, t') = \frac{2}{\pi^s} \sum_{n=1}^{\infty} \frac{\sin(\pi n t) \sin(\pi n t')}{n^{s-\rho}} = \frac{1}{\pi^\rho} K^{s-\rho}(t, t'). \quad (5.2)$$

By Theorem 4.3, samples of $(X(t))_{t \in T} \sim G\mathcal{P}(0, K^s)$ are in the RKHS of this kernel for any such $\rho \geq 2$ because $\sum_{n=1}^{\infty} n^{-\rho} < \infty$ whenever $\rho > 1$. From the identity

$$2 \int_0^1 \sin(\pi n t) \sin(\pi n t') \, dt = \delta_{nm}$$

we see that the Mercer expansion with respect to the Lebesgue measure on $[0, 1]$ of $K^s$ is

$$K^s(t, t') = \sum_{n=1}^{\infty} \lambda_n \psi_n(t) \psi_n(t') \text{ with } \lambda_n = \frac{1}{(\pi n)^s} \text{ and } \psi_n(t) = \sqrt{2} \sin(\pi n t).$$
Small Sample Spaces for Gaussian Processes

Figure 1. Translates (left) and their second derivatives (right) at $t' = 0.3$ of the kernel $K^s$ and its three translates $K_{A_1,\Phi}^s$, $K_{A',\Phi}^s$, and $K_{A,\Phi}^s$ for $s = 4$. The second derivatives have been scaled so as to have $-1$ as their minimum. The scaled kernels were computed by truncating the series after 5,000 terms.

Therefore the $\theta$th power of $H(K^s)$, with $\theta = 1 - \rho/s$, equals $H(K_{A,\Phi}^s)$ as a set. In this case the power RKHS is recovered as an instance of a scaled RKHS. We can also consider logarithmic scalings (recall Section 4.5), which do not correspond to power RKHSs, such as $A = (a_n)^{\infty}_{n=1}$ for $a_n = n \log(n+1)^2$ and $A' = (a'_n)^{\infty}_{n=1}$ for $a'_n = n \log(n+1)$. These yield the scaled kernels

$$K_{A,\Phi}^s(t,t') = \frac{2}{\pi^s} \sum_{n=1}^{\infty} \frac{\sin(\pi nt) \sin(\pi nt')}{n^{s-1}} \log(n+1)^2$$  \hspace{1cm} (5.3)$$

and

$$K_{A',\Phi}^s(t,t') = \frac{2}{\pi^s} \sum_{n=1}^{\infty} \frac{\sin(\pi nt) \sin(\pi nt')}{n^{s-1}} \log(n+1),$$ \hspace{1cm} (5.4)$$

which do not appear to have closed form expressions. Because

$$\sum_{n=1}^{\infty} \frac{1}{n \log(n+1)^2} < \infty \quad \text{but} \quad \sum_{n=1}^{\infty} \frac{1}{n \log(n+1)} = \infty,$$

the samples from $(X(t))_{t \in T} \sim \mathcal{GP}(0, K^s)$ are located in $H(K_{A,\Phi}^s)$ but not in $H(K_{A',\Phi}^s)$.

Translates at $t' = 0.3$ of the kernel $K^s$ in (5.1) and the scaled kernels $K_{A_1,\Phi}^s$ in (5.2), $K_{A,\Phi}^s$ in (5.3) and $K_{A',\Phi}^s$ in (5.4) are plotted in Figure 1 for $s = 4$. Also plotted are the second derivatives of the translates. As noted, the samples from $(X(t))_{t \in T} \sim \mathcal{GP}(0, K^s)$ are in $H(K_{A,\Phi}^s)$ but not in $H(K_{A',\Phi}^s)$ or $H(K_{A_1,\Phi}^s)$. Moreover,

$$H(K^s) \subsetneq H(K_{A_1,\Phi}^s) \subsetneq H(K_{A',\Phi}^s) \subsetneq H(K_{A,\Phi}^s)$$ \hspace{1cm} (5.5)$$

by Proposition 3.3. The second derivative of $K^s(\cdot, t')$ is Lipschitz while those of the scaled kernels are less well-behaved, though still continuous. How the second derivatives behave is indicative of
the inclusions in (5.5): the larger the RKHS, the more severe the non-differentiability of the second derivative at \( t = 0.3 \).

### 5.2. Gaussian Kernel

The ubiquitous Gaussian kernel with a length-scale parameter \( \ell > 0 \) is

\[
K(t, t') = \exp \left( -\frac{(t - t')^2}{2\ell^2} \right) = \exp \left( -\frac{t^2}{2\ell^2} \right) \exp \left( -\frac{t'^2}{2\ell^2} \right) \sum_{n=0}^{\infty} \frac{1}{n!} \frac{(tt')^n}{\ell^{2n}}
\]

and the functions

\[
\phi_n(t) = \frac{1}{\sqrt{n!}} t^n \exp \left( -\frac{t^2}{2\ell^2} \right) \quad \text{for} \ n \geq 0
\]

form an orthonormal basis of its RKHS \([38, 25]\).

The easiest way to proceed is to consider hyperharmonic scalings \( A_\rho = (n^\rho)_{n=0}^{\infty} \) for \( \rho > 0 \) with the convention \( 0^0 = 1 \). Then

\[
K_{A_\rho, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \left( 1 + \sum_{n=1}^{\infty} \frac{n^\rho}{n!} \frac{(tt')^n}{\ell^{2n}} \right).
\]

For \( \rho = 1 \) we get a simple analytic expression for the scaled kernel:

\[
K_{A_1, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \left( 1 + \frac{t t'}{\ell^2} \sum_{n=0}^{\infty} \frac{n^2}{n!} \left( \frac{tt'}{\ell^2} \right)^n \right)
\]

Since the harmonic series \( \sum_{n=1}^{\infty} n^{-1} \) is the prototypical example of an “almost” convergent series, we can informally say that the RKHS of (5.9) is only slightly too small to contain samples of the Gaussian process \((X(t))_{t \in T} \sim \mathcal{G}\mathcal{P}(0, K)\) with covariance kernel (5.6). For a few smaller values of \( \rho \) we compute that

\[
K_{A_2, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \left[ 1 + (a + a^2) \exp(a) \right],
\]

\[
K_{A_3, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \left[ 1 + (a + 3a^2 + a^3) \exp(a) \right],
\]

\[
K_{A_4, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \left[ 1 + (a + 7a^2 + 6a^3 + a^4) \exp(a) \right],
\]

where \( a = t'^2/\ell^2 \). By Theorem 4.3, the RKHSs of the above kernels contain the samples because \( \sum_{n=1}^{\infty} n^{-\rho} < \infty \) if and only if \( \rho > 1 \).

Another example can be constructed using the scaling \( A = (\tau^{2n})_{n=0}^{\infty} \) for some \( \tau > 0 \). Then

\[
K_{A, \Phi}(t, t') = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} \right) \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{\tau^2 tt'}{\ell^2} \right)^n = \exp \left( -\frac{t^2 + (t')^2}{2\ell^2} + \frac{\tau^2 tt'}{\ell^2} \right).
\]
The RKHS of this kernel contains the samples if and only if \( \sum_{n=0}^{\infty} \tau^{-2n} \) converges, which is equivalent to \( \tau > 1 \). Since \( K_{A,\Phi} \) equals the original kernel if \( \tau = 1 \), this class of scalings is not particularly expressive. Because the Mercer eigenvalues of the Gaussian kernel have an exponential decay this example is reminiscent (but more explicit) of the power RKHS construction for the Gaussian kernel in Section 4.4 of Kanagawa et al. [17]. Observe also that with the selection \( \ell^2 = (1 - r^2)/r^2 \) and \( \tau^2 = 1/r \) for \( r \in (0, 1) \) the kernel equals the scaled Mehler kernel

\[
K_M^r(t, t') = \exp \left( -\frac{r^2 (t^2 + (t')^2) - 2rtt'}{2(1 - r^2)} \right) = \sqrt{1 - r^2} \sum_{n=0}^{\infty} \frac{r^n}{n!} H_n(t) H_n(t'),
\]

where \( H_n \) is the \( n \)th probabilists’ Hermite polynomial. The RKHS the Mehler kernel is analysed in [14].

A few of the kernels mentioned above are shown in Figure 2. The RKHSs of the plotted kernels satisfy

\[
H(K) \subsetneq H(K_{A,\Phi}) \subsetneq H(K_{A,\Phi}) \subsetneq H(K_{A,\Phi}) \subsetneq H(K_{A,\Phi})
\]

and the three largest of these contain the samples of \( (X(t))_{t \in T} \sim \mathcal{GP}(0, K) \). All the kernels are qualitatively quite similar, being all infinitely differentiable.

### 5.3. Power Series Kernels

A power series kernel [44, 45] is a kernel of the form

\[
K(t, t') = \sum_{n=0}^{\infty} \frac{w_n}{(n!)^2} (tt')^n
\]  

for positive coefficients \( w_n \) such that the series converges for all \( t, t' \in T \). These kernels do not appear have seen much use in the probabilistic setting, but are useful in functional analysis and approximation theory because, for particular choices of the coefficients, they reproduce important classical function.
spaces. For instance, the selection $w_n = (n!)^2$ yields the Szegő kernel

\[ K(t, t') = \sum_{n=0}^{\infty} (tt')^n = \frac{1}{1 - tt'} \tag{5.12} \]

which converges as long as $T \subset (-1, 1)$ and whose complex extension is the reproducing kernel of the Hardy space $H^2_1$ on the unit disc. The exponential kernel

\[ K(t, t') = \sum_{n=0}^{\infty} \frac{1}{n!} (tt')^n = \exp(tt'), \tag{5.13} \]

obtained by setting $w_n = n!$, is power series kernel that is convergent on the whole real line.

From the expansion in (5.11) we see that the functions

\[ \phi_n(t) = \frac{\sqrt{w_n}}{n!} t^n \quad \text{for } n \geq 0 \]

form an orthonormal basis of $H(K)$ and therefore a scaled kernel is

\[ K_{A, \Phi}(t, t') = \sum_{n=0}^{\infty} \frac{\alpha_n w_n}{(n!)^2} (tt')^n, \]

which itself is a power series kernel. By Theorem 4.3 the samples from $(X(t))_{t \in T} \sim GP(0, K)$ are in the RKHS of a different power series kernel

\[ \bar{K}(t, t') = \sum_{n=0}^{\infty} \frac{w_n}{(n!)^2} (tt')^n \]

if and only if

\[ \sum_{n=0}^{\infty} \frac{w_n}{\bar{w}_n} < \infty. \]

For example, samples from a Gaussian process with the exponential covariance kernel (5.13) are in the RKHS of the Szegő kernel (5.12) if $T \subset (-1, 1)$ because

\[ \sum_{n=0}^{\infty} \frac{w_n}{\bar{w}_n} = \sum_{n=0}^{\infty} \frac{n!}{(n!)^2} < \infty. \]

Its samples are not in the RKHS of the kernel

\[ \bar{K}(t, t') = 1 + \sum_{n=1}^{\infty} \frac{n!}{(n!)^2} (tt')^n = 1 + tt' \exp(tt'), \]

where $w_0 = 1$ and $w_n = n!n$ for $n \geq 1$, because

\[ \sum_{n=0}^{\infty} \frac{w_n}{\bar{w}_n} = 1 + \sum_{n=1}^{\infty} \frac{1}{n} = \infty. \]
6. Application to Maximum Likelihood Estimation

Gaussian processes are often used to model deterministic data-generating functions in, for example, design of computer experiments [32] and probabilistic numerical computation [8, 6]. In applications it is typical that the covariance kernel has parameters, such as the length-scale parameter in (5.6), which are estimated from the data. Maximum likelihood estimation is one of the most popular approaches to estimate the kernel parameters; see, for example, [33, Chapter 3] or [31, Section 5.4.1]. In this section we use Theorem 4.3 to explain the behaviour of the maximum likelihood estimate of the kernel scaling parameter that was recently observed by Xu and Stein [41].

Let $K: T \times T \to \mathbb{R}$ be a positive-definite kernel and $K_\sigma = \sigma^2 K$ a kernel parametrised by a non-negative scaling parameter $\sigma$. Suppose that the data consist of evaluations of a function $f: T \to \mathbb{R}$ at distinct points $t_1, \ldots, t_N \in T$ and that $f$ is modelled as a Gaussian process $(X_\sigma(t))_{t \in T} \sim \mathcal{GP}(0, K_\sigma)$. It is straightforward to compute that the maximum likelihood estimate $\hat{\sigma}_{f,N}$ of $\sigma$ is (see the references above)

$$\hat{\sigma}_{f,N} = \sqrt{\frac{\sum_{i=1}^{N} f_N(t_i) K_{\sigma}^{-1} f_N(t_i)}{N}},$$

where $f_N = (f(t_1), \ldots, f(t_N)) \in \mathbb{R}^N$ is a column vector and $K_N \in \mathbb{R}^{N \times N}$ is the positive-definite covariance matrix with elements $(K_N)_{ij} = K(t_i, t_j)$. Suppose for a moment that $f$ is a zero-mean Gaussian process with covariance $\sigma_0^2 K$. Then

$$\mathbb{E}_f[\sigma^2_{f,N}] = \frac{\mathbb{E}_f[\sum_{i=1}^{N} f_N(t_i) K_{\sigma}^{-1} f_N(t_i)]}{N} = \frac{\mathbb{tr}(\sigma_0^2 K_{\sigma} K_{\sigma}^{-1})}{N} = \frac{\mathbb{tr}(\sigma_0^2 K N)}{N} = \sigma_0^2.$$

This suggests that $\hat{\sigma}_{f,N}$ ought to tend to a constant as $N \to \infty$ if $f$ is a deterministic function which is “akin” to the samples of $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$. Theorem 4.8 can be interpreted as saying that functions of the form $f = \sum_{n=1}^{\infty} f_n \phi_n$ for coefficients $f_n$ which are bounded away from zero and infinity are akin to the samples. For example, the function

$$f(t) = \exp \left(-\frac{t^2}{2\ell^2}\right) \sum_{n=0}^{\infty} \frac{1}{m! n!} t^n$$

can be thought of as a sample from a Gaussian process with the Gaussian covariance kernel in (5.6).

When $H(K)$ is norm-equivalent to the Sobolev space $W_2^{s+d/2}(T)$ on a suitable bounded domain $T \subset \mathbb{R}^d$, Karvonen et al. [18] have used results in the scattered data approximation literature to essentially argue that the maximum likelihood estimate decays to zero if $f$ is too regular to be a sample from $X$ and explodes if $f$ is too irregular in the sense that, assuming the points $t_i$ cover $T$ sufficiently uniformly,

$$\lim_{N \to \infty} \hat{\sigma}_{f,N} = 0 \text{ if } f \in W_2^s(T) \quad \text{and} \quad \lim_{N \to \infty} \hat{\sigma}_{f,N} = \infty \text{ if } f \in W_2^s(T) \setminus W_2^{s-\varepsilon}(T)$$

for any $\varepsilon > 0$ and $r \in (d/2, s - \varepsilon)$. Recall from Section 2 that the samples are contained in $W_2^s(T)$ if and only if $r < s$. This suggests the conjecture that, for any kernels $K_0$, $K_1$, and $K_2$ such that $\mathbb{P}[X \in H(K_0)] = 0$, $\mathbb{P}[X \in H(K_1)] = 1$, and $H(K_2) \subset H(K_1)$, it should hold that

$$\lim_{N \to \infty} \hat{\sigma}_{f,N} = 0 \text{ if } f \in H(K_0) \quad \text{and} \quad \lim_{N \to \infty} \hat{\sigma}_{f,N} = \infty \text{ if } f \in H(K_1) \setminus H(K_2) \quad (6.1)$$
under some assumptions on \( T \) and the points \( t_i \). We now show that the behaviour of \( \hat{\sigma}_{f,N} \) that was observed and conjectured by Xu and Stein [41] when \( K \) is the Gaussian kernel in (5.6) and \( f \) is a monomial agrees with this conjecture if the kernels \( K_0 \), \( K_1 \), and \( K_2 \) are obtained via hyperharmonic scalings.

Let \( K \) be the Gaussian kernel in (5.6). Let \( T = [0, 1] \) and consider the uniform points \( t_i = i/N \) for \( i = 1, \ldots, N \). Xu and Stein [41] used explicit Cholesky decomposition formulae derived in [22] to prove that, for certain positive constants \( C_{\ell,0} \) and \( C_{\ell,1} \),

\[
\hat{\sigma}_{f,N}^2 \sim C_{\ell,0}N^{-1/2} \quad \text{as} \quad N \to \infty
\]

if \( f \equiv 1 \) and

\[
\liminf_{N \to \infty} N^{-1/2} \hat{\sigma}_{f,N}^2 \geq C_{\ell,1}
\]

if \( f(t) = t \). Furthermore, they conjectured that

\[
\hat{\sigma}_{f,N}^2 \sim C_{\ell,p}N^{p-1/2} \quad \text{as} \quad N \to \infty
\]

(6.2) if \( f(t) = t^p \) for any \( p \geq 0 \) and a positive constant \( C_{\ell,p} \). In particular, the conjecture (6.2) implies that \( \hat{\sigma}_{f,N} \to 0 \) if \( f \) is a constant function and \( \hat{\sigma}_{f,N} \to \infty \) if \( f(t) = t^p \) for \( p \in \mathbb{N} \). Xu and Stein [41, p. 142] lacked an intuitive explanation for these findings. We employ a technique similar to that used in [25] to demonstrate that this phenomenon can be explained by the conjecture in (6.1).

Any monomial of degree \( p \in \mathbb{N}_0 \) can be written in terms of the orthonormal basis (5.7) of the Gaussian kernel as

\[
t^p = \exp \left( \frac{t^2}{2\ell^2} \right) t^p \exp \left( - \frac{t^2}{2\ell^2} \right) = \sum_{n=0}^{\infty} \frac{t^{2n+p}}{2^n n!} \exp \left( - \frac{t^2}{2\ell^2} \right) = \sum_{n=0}^{\infty} \frac{(p \sqrt{2n+p})!}{2^n n!} \phi_{2n+p}(t).
\]

(6.3)

Let \( A = (\alpha_n)_{n=1}^{\infty} \) be a \( \Phi \)-scaling of \( H(K) \). If the \( p \)th monomial is to be an element of the scaled RKHS \( H(K_{A,\Phi}) \) it must therefore hold that

\[
\sum_{n=0}^{\infty} \alpha_n^{-1} \frac{(2n+p)!}{2^{2n}(n!)^2} < \infty.
\]

(6.4)

Stirling’s formula gives that, as \( n \to \infty \),

\[
\frac{(2n+p)!}{2^{2n}(n!)^2} \sim \frac{\sqrt{2\pi(2n+p)^2} e^{(2n+p)}}{2\pi 2^{2n+1} n^{2n+1} e^{-2n}} = \frac{1}{\sqrt{2\pi e^p}} \frac{(2n+p)^{2n+p+1/2} e^{-(2n+p)}}{(2n)^{2n}} = \frac{1}{\sqrt{2\pi e^p}} \left( 1 + \frac{p}{2n} \right)^{2n} \left( 2n \right)^{2n+p+1/2} \frac{e^{-(2n+p)}}{n}.
\]

As pointed out by Dette and Zhigljavsky [7, Section 4.1], it appears that the constant \( C_{\ell,p} = \ell^p/(\sqrt{2\pi(p+1/2)}) \) in [41] is erroneous. By truncating the expansion (6.3) after \( N \) terms one arrives, after some relatively straightforward computations, to the conjecture that the constant is \( C_{\ell,p} = 2^p(2^p/(\sqrt{2\pi(p+1/2)}) \).
Since \( \lim_{n \to \infty} (1 + \frac{p}{2n})^{2n} = e^p \), we conclude that
\[
\frac{(2n + p)!}{2^{2n}(n!)^2} \sim \frac{2^p}{\sqrt{\pi}} n^{p-1/2} \quad \text{as} \quad n \to \infty.
\]
Therefore (6.4) holds if and only if \( \sum_{n=1}^{\infty} (2n + p)^{p} n^{p-1/2} < \infty \). If we constraint ourselves to hyperharmonic scalings this is clearly equivalent to \( \alpha_n = n^p \) for \( \rho > p + 1/2 \). But because \( \sum_{n=1}^{\infty} n^{-\rho} < \infty \) if and only if \( \rho > 1 \) it follows that (a) if \( p < 1/2 \), then \( f(t) = t^{p} \) is an element of a hyperharmonic scaled RKHS \( H(K_{A,\Phi}) \) such that \( \mathbb{P}[X \in H(K_{A,\Phi})] = 0 \) and (b) if \( p \geq 1/2 \), then \( f(t) = t^{p} \) is an element of a hyperharmonic scaled RKHS \( H(K_{A,\Phi}) \) if and only if \( \mathbb{P}[X \in H(K_{A,\Phi})] = 1 \). These observations and the conjecture in (6.2) are compatible with our conjecture in (6.1):

- If \( f \) is a constant function, then \( \hat{\sigma}_{f,N} \) tends to zero as \( N \to \infty \) and, for \( A = (n^p)_{n=0}^{\infty} \) and any \( \rho \in (1/2, 1] \), \( \hat{\sigma}_{f,N} \) is an element of the RKHS \( H(K_{0}) = H(K_{A,\Phi}) \) which is of zero measure.
- If \( f(t) = t^{p} \) for \( p \in \mathbb{N} \), then \( \hat{\sigma}_{f,N} \) tends to infinity (or is conjectured to) as \( N \to \infty \). Let \( H(K_{1}) = H(K_{A,\Phi}) \) and \( H(K_{2}) = H(K_{A',\Phi}) \) for hyperharmonic scalings \( A = (n^p)_{n=0}^{\infty} \) and \( A' = (n^p)_{n=0}^{\infty} \), where \( \rho > p + 1/2 \) and \( \sigma \in (1, p + 1/2) \). Then both \( H(K_{1}) \) and \( H(K_{2}) \) are of measure one and \( f \in H(K_{1}) \cap H(K_{2}) \).

In other words, the constant function is too close to \( H(K) \), or too “regular”, to be a sample while higher-order monomials are too far from \( H(K) \), or too “irregular”.

Acknowledgements

This work was supported by the Lloyd’s Register Foundation Programme for Data-Centric Engineering at the Alan Turing Institute, United Kingdom and the Academy of Finland postdoctoral researcher grant #338567 “Scalable, adaptive and reliable probabilistic integration”. I am grateful to Motonobu Kanagawa for his patience in enduring persistent questions over the course of a conference in Sydney in July 2019 during which the main ideas underlying this article were first conceived and to Chris Oates for extensive commentary on the draft at various stages of preparation. Discussions with Anatoly Zhigljavsky led to improvements in Section 6.

References

[1] Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Number 12 in Lecture Notes–Monograph Series. Institute of Mathematical Statistics.
[2] Bogachev, V. I. (1998). *Gaussian Measures*. Number 62 in Mathematical Surveys and Monographs. American Mathematical Society.
[3] Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12:2879–2904.
[4] Cavoretto, R., Fasshauer, G. E., and McCourt, M. (2015). An introduction to the Hilbert-Schmidt SVD using iterated Brownian bridge kernels. *Numerical Algorithms*, 68:393–422.
[5] Cialenco, I., Fasshauer, G. E., and Ye, Q. (2012). Approximation of stochastic partial differential equations by a kernel-based collocation method. *International Journal of Computer Mathematics*, 89(18):2543–2561.
[6] Cockayne, J., Oates, C., Sullivan, T., and Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789.
[7] Dette, H. and Zhigljavsky, A. A. (2021). Reproducing kernel Hilbert spaces, polynomials and the classical moment problems. SIAM/ASA Journal on Uncertainty Quantification, 9(4):1589–1614.
[8] Diaconis, P. (1988). Bayesian numerical analysis. In Statistical decision theory and related topics IV, volume 1, pages 163–175. Springer-Verlag New York.
[9] Driscoll, M. F. (1973). The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 26(4):309–316.
[10] Engl, H. W., Hanke, M., and Neubauer, A. (1996). Regularization of Inverse Problems, volume 375 of Mathematics and Its Applications. Springer.
[11] Flaxman, S., Sejdinovic, D., Cunningham, J., and Filippi, S. (2016). Bayesian learning of kernel embeddings. In Uncertainty in Artificial Intelligence, pages 182–191.
[12] Fortet, R. (1973). Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires. Annales de l’Institut Henri Poincaré, section B, 9(1):41–58.
[13] Gualtierotti, A. F. (2015). Detection of Random Signals in Dependent Gaussian Noise. Springer.
[14] Irregerer, C. and Leobacher, G. (2015). High-dimensional integration on $\mathbb{R}^d$, weighted Hermite spaces, and orthogonal transforms. Journal of Complexity, 31(2):174–205.
[15] Kallianpur, G. (1970). Zero-one laws for Gaussian processes. Transactions of the American Mathematical Society, 149:199–211.
[16] Kallianpur, G. (1971). Abstract Wiener processes and their reproducing kernel Hilbert spaces. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 17:113–123.
[17] Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. arXiv:1807.02582v1.
[18] Karvonen, T., Wynne, G., Tronarp, F., Oates, C. J., and Särkkä, S. (2020). Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. SIAM/ASA Journal on Uncertainty Quantification, 8(3):926–958.
[19] Knopp, K. (1951). Theory and Application of Infinite Series. Blackie & Son, 2nd edition.
[20] Krein, S. G. and Petunin, Yu. I. (1966). Scales of Banach spaces. Russian Mathematical Surveys, 21(2):85–159.
[21] LePage, R. (1973). Subgroups of paths and reproducing kernels. The Annals of Probability, 1(2):345–347.
[22] Loh, W.-L. and Kam, T.-K. (2000). Estimating structured correlation matrices in smooth Gaussian random field models. Annals of Statistics, 28(3):880–904.
[23] Lukić, M. N. (2004). Integrated Gaussian processes and their reproducing kernel Hilbert spaces. In Stochastic Processes and Functional Analysis: A Volume of Recent Advances in Honor of M. M. Rao, number 238 in Lecture Notes in Pure and Applied Mathematics, pages 241–263. Marcel Dekker.
[24] Lukić, M. N. and Beder, J. H. (2001). Stochastic processes with sample paths in reproducing kernel Hilbert spaces. Transactions of the American Mathematical Society, 353(10):3945–3969.
[25] Minh, H. Q. (2010). Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. Constructive Approximation, 32(2):307–338.
[26] Mörters, P. and Peres, Y. (2010). Brownian Motion. Number 30 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
[27] Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I. Technical Report 23, Applied Mathematics and Statistics Laboratory, Stanford University.
[28] Parzen, E. (1963). Probability density functionals and reproducing kernel Hilbert spaces. In Proceedings of the Symposium on Time Series Analysis, pages 155–169. John Wiley & Sons.
[29] Paulsen, V. I. and Raghupathi, M. (2016). An Introduction to the Theory of Reproducing Kernel Hilbert Spaces. Number 152 in Cambridge Studies in Advanced Mathematics. Cambridge University Press.
[30] Rakotomamonjy, A. and Canu, S. (2005). Frames, reproducing kernels, regularization and learning. 
Journal of Machine Learning Research, 6:1485–1515.
[31] Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian Processes for Machine Learning.
Adaptive Computation and Machine Learning. MIT Press.
[32] Sacks, J., Welch, W. J., M. T. J., and Wynn, H. P. (1989). Design and analysis of computer 
experiments. Statistical Science, 4(4):409–435.
[33] Santner, T. J., Williams, B. J., and Notz, W. I. (2003). The Design and Analysis of Computer 
Experiments. Springer Series in Statistics. Springer.
[34] Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. 
Stochastic Processes and their Applications, 120(10):1879–1897.
[35] Stein, M. L. (1999). Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in 
Statistics. Springer.
[36] Steinwart, I. (2019). Convergence types and rates in generic Karhunen-Loève expansions with 
applications to sample path properties. Potential Analysis, 51:361–395.
[37] Steinwart, I. and Christmann, A. (2008). Support Vector Machines. Information Science and 
Statistics. Springer.
[38] Steinwart, I., Hush, D., and Scovel, C. (2006). An explicit description of the reproducing kernel 
Hilbert spaces of Gaussian RBF kernels. IEEE Transactions on Information Theory, 52(10):4635–
4643.
[39] Steinwart, I. and Scovel, C. (2012). Mercer’s theorem on general domains: On the interaction 
between measures, kernels, and RKHSs. Constructive Approximation, 35:363–417.
[40] van der Vaart, A. and van Zanten, H. (2011). Information rates of nonparametric Gaussian process 
methods. Journal of Machine Learning Research, 12(6):2095–2119.
[41] Xu, W. and Stein, M. L. (2017). Maximum likelihood estimation for a smooth Gaussian random 
field model. SIAM/ASA Journal on Uncertainty Quantification, 5(1):138–175.
[42] Xu, Y. and Zhang, H. (2009). Refinement of reproducing kernels. Journal of Machine Learning 
Research, 10:107–140.
[43] Zhang, H. and Zhao, L. (2013). On the inclusion relation of reproducing kernel Hilbert spaces. 
Analysis and Applications, 11(2):1350014.
[44] Zwicknagl, B. (2009). Power series kernels. Constructive Approximation, 29(1):61–84.
[45] Zwicknagl, B. and Schaback, R. (2013). Interpolation and approximation in Taylor spaces. Journal 
of Approximation Theory, 171:65–83.