Small sample spaces for Gaussian processes

TONI KARVONEN[1,2,a]

[1]*The Alan Turing Institute, 96 Euston Road, London NW1 2DB, United Kingdom*
[2]*University of Helsinki, Department of Mathematics and Statistics, Pietari Kalmin katu 5, 00560 Helsinki, Finland,* [a]*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 σ-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,

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

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 $\mathbb{P}[X \in H(R_1) \setminus H(R_2)] = 1$ 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 [11,40], emulation and approximation [18,41], 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 [31,35], 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 R such that $R \gg K$. For any orthonormal basis $\Phi = (\phi_n)_{n=1}^\infty$ of $H(K)$ the kernel K 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*

$$K_{A,\Phi}(t,t') = \sum_{n=1}^\infty \alpha_n \phi_n(t)\phi_n(t'). \tag{1.1}$$

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]. We call the sequence A a Φ-*scaling* of $H(K)$. If $\alpha_n \to \infty$ as $n \to \infty$, the corresponding *scaled RKHS*[1], $H(K_{A,\Phi})$, 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 Φ-scaling of $H(K)$. If $K_{A,\Phi}$ is continuous and $d_K(t,t') = \|K(\cdot,t) - K(\cdot,t')\|_K$ is a metric on T, then either*

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

In Section 5, we use this result to study sample path properties of Gaussian processes defined by infinitely smooth kernels. These appear to be the first sufficiently descriptive results of their kind. An example of an infinitely smooth kernel that we consider is the univariate (i.e., $T \subset \mathbb{R}$) Gaussian kernel $K(t,t') = \exp(-(t - t')^2/(2\ell^2))$ with length-scale $\ell > 0$ for which we explicitly construct several scaled kernels R whose RKHSs are "small" but still contain the samples of $(X(t))_{t\in T} \sim \mathcal{GP}(0,K)$. In Section 6, Theorem 4.3 is applied to provide an intuitive explanation for a conjecture by Xu and Stein [41] on asymptotic behaviour of the maximum likelihood estimate of the scaling parameter of the Gaussian kernel when the data are generated by a monomial function on a uniform grid.

Secondly, we use Theorem 4.3 to construct a "small" set which "almost" contains the samples. This *sample support set* is distinct from the traditional topological support of a Gaussian measure; see the discussion at the end of Section 2. Let $C(T)$ denote the set of continuous function on T.

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 Φ-scaling $A = (\alpha_n)_{n=1}^\infty$ such that $\sum_{n=1}^\infty \alpha_n^{-1} < \infty$ and $K_{A,\Phi}$ is continuous. Let $\mathcal{S}(\mathfrak{R})$ be the σ-algebra generated by the collection of scaled RKHSs consisting of continuous functions and $S_\mathfrak{R}(K)$ the largest subset of $C(T)$ that is contained in every $H \in \mathcal{S}(\mathfrak{R})$ such that $\mathbb{P}[X \in H] = 1$. Suppose that $d_K(t,t') = \|K(\cdot,t) - K(\cdot,t')\|_K$ is a metric on T. Then $S_\mathfrak{R}(K)$ consists precisely of the functions $f = \sum_{n=1}^\infty f_n \phi_n$ such that*

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

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

---

[1]These spaces are not to be confused with classical Hilbert scales defined via powers of a strictly positive self-adjoint operator; see [10, Section 8.4] and [20].

The set $S_\Re(K)$ may fail to be measurable. The latter part of the above result is therefore important in demonstrating that it is possible to construct sets which are arbitrarily close to $S_\Re(K)$ and contain the samples using countably many elementary set operations of scaled RKHSs. At its core this is a manifestation of the classical result that there is no 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 ζ_n are independent standard normal random variables. The series converges in $L^2(\mathbb{P})$, but if almost all samples of X are continuous, convergence is also uniform on T with probability one [1, Theorem 3.8]. Because $\|X(t)\|_K^2 = \sum_{n=1}^\infty \zeta_n^2$ and $\mathbb{E}[\zeta_n^2] = 1$ for every n, the Karhunen–Loève expansion suggests, somewhat informally, that the samples are functions $f = \sum_{n=1}^\infty f_n \phi_n$ for which the sequence $(f_n^2)_{n=1}^\infty$ satisfies (1.2).

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 $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ 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 $\tilde{\mathcal{B}}$ the σ-algebra generated by cylinder sets of the form $\{f \in \mathbb{R}^T : (f(t_1), \dots, f(t_n)) \in B^n\}$ for any $n \in \mathbb{N}$ and any Borel set $B^n \subset \mathbb{R}^n$. Let $\Phi_X(\omega) = X(\cdot, \omega)$. Then $\tilde{\mu}_X = \mathbb{P} \circ \Phi_X^{-1}$ is the law of X on the measurable space $(\mathbb{R}^T, \tilde{\mathcal{B}})$. Consequently, $\mathbb{P}[X \in H] = \tilde{\mu}_X(\{\omega \in \Omega : X(\cdot, \omega) \in H\})$ for $H \in \tilde{\mathcal{B}}$. Let $(\mathbb{R}^T, \tilde{\mathcal{B}}_0, \tilde{\mu}_{X,0})$ be the completion of $(\mathbb{R}^T, \tilde{\mathcal{B}}, \tilde{\mu}_X)$ and $R \colon T \times T \to \mathbb{R}$ a positive-semidefinite kernel and $H(R)$ its RKHS. The following facts are known about the measurability of $H(K)$ and $H(R)$:

- In general, $H(R) \notin \tilde{\mathcal{B}}$. For example, if T is equipped with a topology and R is continuous on $T \times T$, then $H(R) \subset C(T)$. However, no non-empty subset of $C(T)$ can be an element of $\tilde{\mathcal{B}}$.
- LePage [21, p. 347] has proved that $H(K) \in \tilde{\mathcal{B}}_0$ and $\tilde{\mu}_{X,0}(H(K)) = 0$, a claim which originates with Parzen [27, 28]. A version which requires separability and continuity appears in [16].
- If the RKHS $H(R)$ is infinite-dimensional and $R \gg K$, then $H(R) \in \tilde{\mathcal{B}}_0$ and $\tilde{\mu}_{X,0}(H(R)) = 0$. 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 $G \in \tilde{\mathcal{B}}_0$, then either $\tilde{\mu}_{X,0}(G) = 0$ or $\tilde{\mu}_{X,0}(G) = 1$. Furthermore, $\tilde{\mu}_{X,0}(G) = 1$ implies that $H(K) \subset G$. This is a general version of the 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, that the kernel K is continuous, and that almost all samples of $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ are continuous, which is to say that the $(\tilde{\mathcal{B}}_0, \tilde{\mu}_{X,0})$-outer measure of $C(T)$ is one. Define the probability space $(C(T), \mathcal{B}, \mu_X)$ as

$$\mathcal{B} = C(T) \cap \tilde{\mathcal{B}}_0 \quad \text{and} \quad \mu_X(C(T) \cap H) = \tilde{\mu}_{X,0}(H) \text{ for } H \in \tilde{\mathcal{B}}_0. \tag{1.3}$$

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 for a continuous and positive-semidefinite R; 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 $a_n \lesssim b_n$ if there is $C > 0$ such that $a_n \leq C b_n$ for all sufficiently large n. If both $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold, 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 $F \subsetneq G$ to indicate that F is a proper subset of G. A kernel $R \colon T \times T \to \mathbb{R}$ is *positive-semidefinite* if

$$\sum_{i=1}^N \sum_{j=1}^N a_i a_j R(t_i, t_j) \geq 0 \tag{1.4}$$

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) $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ is a zero-mean Gaussian process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with *continuous paths*. The law μ_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]. We do not pursue these generalisations.

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 [23], convolution kernels in [5] and [11, Section 3.1], and, most importantly, *powers of RKHSs* [39] in [17, Section 4] and [36]. Namely, let T be a compact metric space, K a continuous kernel on $T \times T$, and ν a finite and strictly positive Borel measure on T. Then the integral operator T_ν, defined for $f \in L^2(\nu)$ via

$$(T_\nu f)(t) = \int_T K(t, t') f(t') \, d\nu(t'), \tag{2.1}$$

has decreasing and positive eigenvalues $(\lambda_n)_{n=1}^\infty$, which vanish as $n \to \infty$, and eigenfunctions $(\psi_n)_{n=1}^\infty$ in $H(K)$ such that $(\sqrt{\lambda_n} \psi_n)_{n=1}^\infty$ 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 θth power of K is defined as

$$K^{(\theta)}(t, t') = \sum_{n=1}^\infty \lambda_n^\theta \psi_n(t) \psi_n(t'). \tag{2.2}$$

The series (2.2) converges if $\sum_{n=1}^\infty \lambda_n^\theta \psi_n(t)^2 < \infty$ for all $t \in T$. Furthermore, $H(K^{(\theta_2)}) \subsetneq H(K^{(\theta_1)})$ if $\theta_1 < \theta_2$ and $\mathbb{P}[X \in H(K^{(\theta)})] = 1$ if and only if $\sum_{n=1}^\infty \lambda_n^{1-\theta} < \infty$ [36, Theorem 5.2]. When it comes to

---

[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].

sample properties, the power kernel construction has two significant downsides: (i) The measure ν 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 $H(K^{(\theta)})$ 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 $H(K) = H(K^{(\theta=1)})$ 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 polynomial decay n^{-a} for $a > 0$ (e.g., Matérn kernels) than when the kernel is infinitely smooth with at least exponential eigenvalue decay e^{-bn} for $b > 0$ (e.g., Gaussian): the change the decay condition $\sum_{n=1}^{\infty} f_n^2 < \infty$ for the coefficients $(f_n)_{n=1}^{\infty}$ to $\sum_{n=1}^{\infty} f_n^2 n^{-a(1-\theta)} < \infty$ is arguably less substantial than that from $\sum_{n=1}^{\infty} f_n^2 < \infty$ to $\sum_{n=1}^{\infty} f_n^2 e^{-b(1-\theta)n} < \infty$. Indeed, as pointed out by Kanagawa et al. [17, Section 4.4], 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 because $H(K^{(\theta=1)}) = H(K)$ does not contain the samples. The relationship between powers of RKHSs and scaled RKHSs is discussed in more detail 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 [36, pp. 369–370], given here in a somewhat applied form and without some technicalities, that the samples are contained in the set

$$\left(\bigcap_{r<s} W_2^r(T) \right) \setminus W_2^s(T) \tag{2.3}$$

with probability one if $H(K)$ is norm-equivalent to the fractional Sobolev space $W_2^{s+d/2}(T)$ 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 $W_2^s(T) = \cup_{r \geq s} W_2^r(T)$, 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 $W_2^1([0,1])$ with zero boundary condition at the origin, have regularity $1/2$ in the sense that they are almost surely α-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 $|X(t,\omega) - X(t',\omega)| \leq C\sqrt{h \log(1/h)}$ when $h = |t - t'|$ is sufficiently small. Since the Sobolev space $W_2^s(T)$ 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

$$\int_{\mathbb{R}^d} \left(1 + \|\xi\|^2 \right)^s \left| \widehat{f_e}(\xi) \right|^2 \mathrm{d}\xi < \infty, \tag{2.4}$$

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 $W_2^s(T)$ and a proper subset of $W_2^r(T)$ for every $r < s$ and hence a proper subset of the set in (2.3). Some results and discussion in [18, Section 4.2] and [34] 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*, $\mathrm{supp}_{C(T)}(\mu_X)$, of the measure μ_X is the set

$$\mathrm{supp}_{C(T)}(\mu_X) = \{f \in C(T) : \mu_X(B(f,r)) > 0 \text{ for all } r > 0\},$$

where $B(f,r)$ is the f-centered r-ball in $(C(T), \|\cdot\|_\infty)$. It is a classical result [16, Theorem 3] that $\operatorname{supp}_{C(T)}(\mu_X) = \overline{H(K)}$, where $\overline{H(K)}$ is the closure of $H(K)$ in $(C(T), \|\cdot\|_\infty)$. In other words, the topological support of μ_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 $\operatorname{supp}_{C(T)}(\mu_X) = \overline{H(K)} = C(T)$ 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 [2, Section 3.6] for further results on general topological supports of Gaussian measures.

3. Scaled reproducing kernel Hilbert spaces

For any orthonormal basis $\Phi = (\phi_n)_{n=1}^\infty$ of $H(K)$ the kernel has the pointwise convergent expansion $K(t,t') = \sum_{n=1}^\infty \phi_n(t)\phi_n(t')$ for all $t, t' \in T$. By the standard characterisation of a separable Hilbert space, the RKHS consists of precisely those functions $f \colon T \to \mathbb{R}$ that admit an expansion $f = \sum_{n=1}^\infty f_n \phi_n$ for coefficients such that $\sum_{n=1}^\infty f_n^2 < \infty$. The Cauchy–Schwarz inequality ensures that 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 $\langle f, g \rangle_K = \sum_{n=1}^\infty f_n g_n$.

Definition 3.1 (Scaled kernel and RKHS). We say that a positive sequence $A = (\alpha_n)_{n=1}^\infty$ is a Φ-*scaling* of $H(K)$ if $\sum_{n=1}^\infty \alpha_n \phi_n(t)^2 < \infty$ for every $t \in T$. The kernel

$$K_{A,\Phi}(t,t') = \sum_{n=1}^\infty \alpha_n \phi_n(t)\phi_n(t') \tag{3.1}$$

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* Φ-*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*

$$H(K_{A,\Phi}) = \left\{ f = \sum_{n=1}^\infty f_n \phi_n \; : \; \|f\|_{K_{A,\Phi}}^2 = \sum_{n=1}^\infty \frac{f_n^2}{\alpha_n} < \infty \right\}, \tag{3.2}$$

where convergence is pointwise, and for any $f = \sum_{n=1}^\infty f_n \phi_n$ *and* $g = \sum_{n=1}^\infty g_n \phi_n$ *in* $H(K_{A,\Phi})$ *its inner product is*

$$\langle f, g \rangle_{K_{A,\Phi}} = \sum_{n=1}^\infty \frac{f_n g_n}{\alpha_n}. \tag{3.3}$$

Proof. By the Cauchy–Schwarz inequality and $\sum_{n=1}^\infty \alpha_n \phi_n(t)^2 < \infty$ for every $t \in T$,

$$\sum_{n=1}^\infty |\alpha_n \phi_n(t)\phi_n(t')| \leq \left(\sum_{n=1}^\infty \alpha_n \phi_n(t)^2 \right)^{1/2} \left(\sum_{n=1}^\infty \alpha_n \phi_n(t')^2 \right)^{1/2} < \infty$$

for any $t, t' \in T$. This proves that the scaled kernel in (3.1) 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 α_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 (3.2) and (3.3) 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]. $\square$

A scaled RKHS depends on the ordering of the orthonormal basis of $H(K)$ used to construct it. For example, let $\Phi = (\phi_n)_{n=1}^{\infty}$ be an orthonormal basis of $H(K)$ and suppose that $\alpha_n = n$ defines a Φ-scaling of $H(K)$. Define another ordered orthonormal basis $\Psi = (\psi_n)_{n=1}^{\infty}$ by setting $\psi_{2n+1} = \phi_{2^n}$ for $n \geq 0$ and interleaving the remaining ϕ_n to produce $\Psi = (\phi_1, \phi_3, \phi_2, \phi_5, \phi_4, \phi_6, \phi_8, \phi_7, \phi_{16}, \ldots)$. The function $f = \sum_{n=0}^{\infty} \phi_{2^n} =: \sum_{n=1}^{\infty} f_{\Phi,n} \phi_n$ is in $H(K_{A,\Phi})$ because

$$\|f\|_{K_{A,\Phi}}^2 = \sum_{n=1}^{\infty} \frac{f_{\Phi,n}^2}{\alpha_n} = \sum_{n=0}^{\infty} \frac{1}{2^n} < \infty$$

but not in $H(K_{A,\Psi})$ because $f = \sum_{n=0}^{\infty} \phi_{2^n} = \sum_{n=0}^{\infty} \psi_{2n+1} =: \sum_{n=1}^{\infty} f_{\Psi,n} \psi_n$ and therefore

$$\|f\|_{K_{A,\Psi}}^2 = \sum_{n=1}^{\infty} \frac{f_{\Psi,n}^2}{\alpha_n} = \sum_{n=0}^{\infty} \frac{1}{2n+1} = \infty.$$

In practice, the orthonormal basis usually has a natural ordering. For instance, the decreasing eigenvalues specify an 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 α_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 Φ-scalings of $H(K)$. Then $H(K_{B,\Phi}) \subset H(K_{A,\Phi})$ if and only if $\beta_n \preceq \alpha_n$. In particular, $H(K) \subset H(K_{A,\Phi})$ if and only if $\inf_{n \geq 1} \alpha_n > 0$.*

Proof. If $\beta_n \preceq \alpha_n$, then for any $f = \sum_{n=1}^{\infty} f_n \phi_n \in H(K_{B,\Phi})$ we have

$$\|f\|_{K_{A,\Phi}}^2 = \sum_{n=1}^{\infty} \frac{f_n^2}{\alpha_n} = \sum_{n=1}^{\infty} \frac{\beta_n}{\alpha_n} \frac{f_n^2}{\beta_n} \leq \|f\|_{K_{B,\Phi}}^2 \sup_{n \geq 1} \frac{\beta_n}{\alpha_n} < \infty. \tag{3.4}$$

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 $\alpha_{n_m}^{-1} \beta_{n_m} \geq 2^m$. Then $f = \sum_{m=1}^\infty 2^{-m/2} \sqrt{\beta_{n_m}} \phi_{n_m} \in H(K_{B,\Phi}) \setminus H(K_{A,\Phi})$ since

$$\|f\|_{K_{B,\Phi}}^2 = \sum_{m=1}^\infty \left(\frac{\sqrt{\beta_{n_m}}}{2^{m/2}} \right)^2 \frac{1}{\beta_{n_m}} = \sum_{m=1}^\infty 2^{-m} = 1 \quad \text{but} \quad \|f\|_{K_{A,\Phi}}^2 = \sum_{m=1}^\infty 2^{-m} \frac{\beta_{n_m}}{\alpha_{n_m}} \geq \sum_{m=1}^\infty 1 = \infty,$$

which contradicts the assumption that $H(K_{B,\Phi}) \subset H(K_{A,\Phi})$. Thus $\sup_{n \geq 1} \alpha_n^{-1} \beta_n < \infty$. 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 $C_1 \|f\|_F \leq \|f\|_G \leq C_2 \|f\|_F$ for all $f \in F$. From (3.4) it follows that $H(K_{A,\Phi})$ and $H(K_{B,\Phi})$ are norm-equivalent if and only 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 Φ-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 Φ-scalings of $H(K)$. Then $H(K_{B,\Phi}) \subsetneq H(K_{A,\Phi})$ if and only if $\sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty$ and $\beta_n \lesssim \alpha_n$.*

Proof. Assume first that $\sup_{n \geq 1} \alpha_n \beta_n^{-1} = \infty$ and $\beta_n \lesssim \alpha_n$. Since $\beta_n \lesssim \alpha_n$, Proposition 3.3 yields $H(K_{B,\Phi}) \subset H(K_{A,\Phi})$. Thus $H(K_{B,\Phi})$ 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 $\alpha_n \lesssim \beta_n$, 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 $H(K_{B,\Phi}) \subsetneq H(K_{A,\Phi})$. Then $\beta_n \lesssim \alpha_n$ by Proposition 3.3. 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 $\alpha_n \lesssim \beta_n$. But by Proposition 3.3 this would imply that $H(K_{A,\Phi}) \subset H(K_{B,\Phi})$, which would contradict the assumption that $H(K_{B,\Phi})$ is a proper subset of $H(K_{A,\Phi})$. This completes the proof. □

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, $(\|\phi_n\|_R^{-1} \phi_n)_{n=1}^\infty$ is an orthonormal basis of $H(R)$ and consequently $H(R) = H(K_{A,\Phi})$ for the scaling with $\alpha_n = \|\phi_n\|_R^{-2}$.

We conclude this section by demonstrating that scaled RKHSs generalise powers of RKHSs. We say that a Φ-scaling $A_\rho = (\alpha_n)_{n=1}^\infty$ of $H(K)$ is *ρ-hyperharmonic* if $\alpha_n = n^\rho$ for some $\rho \geq 0$. The RKHS $H(K_{A_\rho,\Phi})$ is a *ρ-hyperharmonic scaled RKHS*. Recall from Section 2 that if T is a compact metric space, K is continuous on $T \times T$, and ν is a finite and strictly positive Borel measure on T, then the integral operator in (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 θth power of K and its RKHS $H(K^{(\theta)})$ the θth 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 $K_{A,\Psi}$ for the scaling $A = (\lambda_n^{\theta-1})_{n=1}^\infty$ because

$$K_{A,\Psi}(t,t') = \sum_{n=1}^\infty \lambda_n^{\theta-1} \lambda_n \psi_n(t) \psi_n(t') = \sum_{n=1}^\infty \lambda_n^\theta \psi_n(t) \psi_n(t') = K^{(\theta)}(t,t').$$

Polynomially decaying eigenvalues ($\lambda_n \asymp n^{-p}$ for some $p > 0$) are an important special case. For example, this holds with $p = -2s/d$ if $T \subset \mathbb{R}^d$ and $H(K)$ is norm-equivalent to the Sobolev space $W_2^s(T)$ for $s > d/2$ [36, p. 370]. If $\lambda_n \asymp n^{-p}$ and $\rho = p(1 - \theta)$, the ρ-hyperharmonic scaled RKHS is norm-equivalent to the power RKHS $H(K^{(\theta)})$ by Corollary 3.4 since $n^\rho \lambda_n \asymp n^{-\theta p} \asymp \lambda_n^\theta$.

4. Sample path properties

This section contains the main results of the article. First, we consider a specialisation to scaled RKHSs of a theorem originally proved by Driscoll [9] and later generalised by Lukíc and Beder [24]. Then we define general sample support sets and characterise them for σ-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 $L\colon H(R) \to H(K)$, called the *dominance operator*, whose range is contained in $H(K)$ and which satisfies $\langle f,g \rangle_R = \langle Lf,g \rangle_K$ for all $f \in H(R)$ and $g \in H(K)$. 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

$$\mathrm{tr}(L) = \sum_{n=1}^\infty \langle L\psi_n, \psi_n \rangle_R < \infty \tag{4.1}$$

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 only positive-semidefinite because $R(t,0) = 0$ for every $t \in T$ but nevertheless yields a metric because $d_R(t,t') = \sqrt{t - 2\min\{t,t'\} + t'}$ vanishes if and only if $t = t'$. See [24, Section 4] for more properties of d_R. Injectivity of the mapping $t \mapsto R(\cdot,t)$ is equivalent to d_R being a metric.

By 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 the 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*

$$\mathbb{P}\big[X \in H(R)\big] = 0 \;\; and \;\; R \not\gg K \quad or \quad \mathbb{P}\big[X \in H(R)\big] = 1 \;\; and \;\; R \gg K.$$

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 of the samples, which one of our standing assumptions, implies their d_R-continuity. □

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

---

[3] A change of basis shows that $\mathrm{tr}(L)$ does not depend on the orthonormal basis.

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 Φ-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 α_n are positive, we conclude that $d_R(t,t') = 0$ if and only if $t = t'$. $\qquad\square$

Theorem 4.3. *Let $\Phi = (\phi_n)_{n=1}^\infty$ be an orthonormal basis of $H(K)$, $A = (\alpha_n)_{n=1}^\infty$ a Φ-scaling of $H(K)$, and $(X(t))_{t \in T} \sim \mathcal{GP}(0,K)$. If $K_{A,\Phi}$ is continuous and d_K is a metric, then either*

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

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 dominance operator $L\colon H(K_{A,\Phi}) \to H(K)$ is given by $Lf = \sum_{n=1}^\infty f_n \alpha_n^{-1} \phi_n$ for any $f = \sum_{n=1}^\infty f_n \phi_n \in H(K_{A,\Phi})$. Because $(\sqrt{\alpha_n}\phi_n)_{n=1}^\infty$ is an orthonormal basis of $H(K_{A,\Phi})$ and $L(\sqrt{\alpha_n}\phi_n) = 1/\sqrt{\alpha_n}$, the nuclear dominance condition (4.1) is

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

and the claim follows from Theorem 4.1 since Lemma 4.2 guarantees that d_R for $R = K_{A,\Phi}$ is a metric. Assume then that $H(K) \not\subset H(K_{A,\Phi})$. It is trivial that $K_{A,\Phi} \not\gg K$. Thus $\mathbb{P}[X \in H(K_{A,\Phi})] = 0$. 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, which is a contradiction. Therefore $\sum_{n=1}^\infty \alpha_n^{-1} = \infty$. $\qquad\square$

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 $\mathfrak{R}$ be a collection of continuous kernels R on T for which d_R is a metric and $H(\mathfrak{R})$ the corresponding set of RKHSs. Every element of $H(\mathfrak{R})$ is a subset of $C(T)$. By the generalised Driscoll's theorem each element of $H(\mathfrak{R})$ has μ_X-measure one or zero, depending on the nuclear dominance condition. Define the disjoint sets

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

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

Definition 4.4 (Sample support set). Let $\mathcal{S}(\mathfrak{R}) = \sigma(H(\mathfrak{R}))$ be the σ-algebra generated by $H(\mathfrak{R})$. The *sample support set*, $S_\mathfrak{R}(K)$, of the Gaussian process $(X(t))_{t \in T} \sim \mathcal{GP}(0,K)$ with respect to $\mathfrak{R}$ is the largest subset of $C(T)$ such that $S_\mathfrak{R}(K) \subset H$ for every $H \in \mathcal{S}(\mathfrak{R})$ such that $\mu_X(H) = 1$.

Proposition 4.5. *It holds that*

$$S_\mathfrak{R}(K) = \bigcap_{R_1 \in \mathfrak{R}_1(K)} H(R_1) \setminus \bigcup_{R_0 \in \mathfrak{R}_0(K)} H(R_0). \tag{4.2}$$

Proof. Suppose that there is $f \in S_\Re(K)$ which is not contained in the set on the right-hand side of (4.2). That is, we have either $f \notin \cap_{R_1 \in \Re_1(K)} H(R_1)$ or $f \in \cup_{R_0 \in \Re_0(K)} H(R_0)$. In the former case there is $R_1 \in \Re_1(K)$ such that $f \notin H(R_1)$. But because $\mu_X(H(R_1)) = 1$ and $S_\Re(K) \subset H(R_1)$ by definition, this violates the assumption that $f \in S_\Re(K)$. In the latter case there is $R_0 \in \Re_0(K)$ such that $f \in H(R_0)$. As $\mu_X(H(R_0)) = 0$, we have for any $R_1 \in \Re_1$ 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 $f \in S_\Re(K)$ is again violated and we conclude that $S_\Re(K) \subset \cap_{R_1 \in \Re_1(K)} H(R_1) \setminus \cup_{R_0 \in \Re_0(K)} H(R_0)$.

Since all elements of $H(\Re)$ are either of measure zero or one, so are those of $\mathcal{S}(\Re)$. It is therefore clear that $\cap_{R_1 \in \Re_1(K)} H(R_1) \setminus \cup_{R_0 \in \Re_0(K)} H(R_0)$ is contained in every $H \in \mathcal{S}(\Re)$ such that $\mu_X(H) = 1$. Consequently, $\cap_{R_1 \in \Re_1(K)} H(R_1) \setminus \cup_{R_0 \in \Re_0(K)} H(R_0) \subset S_\Re(K)$. This concludes the proof. $\qquad\square$

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 $R \in \Re$. The larger $\Re$ is, the more precisely $S_\Re(K)$ describes the samples of X. But there is an important caveat. If $\Re$ is countable, the sample support set is in the σ-algebra $\mathcal{B}$, defined in (1.3), and has μ_X-measure one. However, when $\Re$ is uncountable and does not contain countable subsets $\Re_1'(K) \subset \Re_1(K)$ and $\Re_0'(K) \subset \Re_0(K)$ such that

$$\bigcap_{R_1 \in \Re_1(K)} H(R_1) = \bigcap_{R_1 \in \Re_1'(K)} H(R_1) \quad \text{and} \quad \bigcup_{R_0 \in \Re_0(K)} H(R_0) = \bigcup_{R_0 \in \Re_0'(K)} H(R_0),$$

it cannot be easily determined if $S_\Re(K)$ is an element of $\mathcal{B}$.

We are mainly interested in sample support sets with respect to $\Re$ which consist 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 an eventuality in mind we have introduced the more general concept of a sample support set. If $\Re$ 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 Σ be a collection of non-negative sequences. A non-negative sequence $(a_n)_{n=1}^\infty$ is said to be Σ-*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 properties of approximately constant sequences: (i) If $(a_n)_{n=1}^\infty$ and $(a_n')_{n=1}^\infty$ are two Σ-approximately constant sequences, then so is their sum. (ii) The larger Σ is, the fewer Σ-approximately sequences there are. That is, if Σ_1 and Σ_2 are two collections of non-negative sequences such that $\Sigma_1 \subset \Sigma_2$, then a non-negative sequence is Σ_1-approximately constant if it is Σ_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 Σ-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 Σ_Φ a collection of Φ-scalings of $H(K)$ such that the corresponding scaled kernels are continuous. Suppose that d_K is a metric and let $\Re = \{K_{A,\Phi} : A \in \Sigma_\Phi\}$. Then $S_\Re(K) = \mathcal{R}(\Sigma_\Phi, \Phi)$.*

Proof. Note first that, by Lemma 4.2, d_R is a metric for every $R \in \Re$. Because every scaling of $H(K)$ has an orthonormal basis that is a scaled version of $(\phi_n)_{n=1}^\infty$, every $f \in S_\Re(K)$ 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, $K_{A,\Phi} \in \Re_1(K)$ if $A \in \Sigma_1(K)$ and $K_{A,\Phi} \in \Re_0(K)$ if $A \in \Sigma_0(K)$. Because, by definition, $S_\Re(K) \subset H(K_{A,\Phi})$ for any $A \in \Sigma_1(K)$ and $S_\Re(K) \cap H(K_{A,\Phi}) = \varnothing$ for any $A \in \Sigma_0(K)$ it follows that for every $f \in S_\Re(K)$ and any $(\alpha_n)_{n=1}^\infty \in \Sigma_\Phi$ we have

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

That is, $(f_n^2)_{n=1}^\infty$ is a Σ_Φ-approximately constant sequence and thus $S_\Re(K) \subset \mathcal{R}(\Sigma_\Phi, \Phi)$. 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 $f \in S_\Re(K)$ and thus $S_\Re(K) = \mathcal{R}(\Sigma_\Phi, \Phi)$. □

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

4.3. Sample support sets for scaled RKHSs

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

$$\liminf_{n\to\infty} a_n > 0 \quad \text{ and } \quad \sup_{n \geq 1} a_n < \infty. \tag{4.3}$$

For suppose that there existed a Σ-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 $(a_{n_m})_{m=1}^\infty$ such that $a_{n_m} \leq 2^{-m}$ 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 $b_{n_m} = 1$ for $m \in \mathbb{N}$. Then $\sum_{n=1}^\infty b_n$ diverges but

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

which contradicts the assumption that $(a_n)_{n=1}^\infty$ is a Σ-approximately constant sequence. A similar argument (with $a_{n_m} \geq 2^m$, $b_n = 2^{-n}$, and $b_{n_m} = 2^{-m}$) shows the second condition in (4.3) cannot be violated either; thus every Σ-approximately constant sequence satisfies (4.3). A sequence satisfying (4.3) is trivially Σ-approximately constant because the conditions imply the existence of constants $0 < c_1 \leq c_2$ 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 full proof is more complicated than the above argument as 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* Φ-*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* Σ_Φ *be the collection of* Φ-*scalings of* $H(K)$ *such that the corresponding scaled kernels are continuous. Suppose that* d_K *is a metric and let* $\Re = \{K_{A,\Phi} : A \in \Sigma_\Phi\}$. *Then* $S_\Re(K)$ *is non-empty and consists precisely of the functions* $f = \sum_{n=1}^\infty f_n \phi_n$ *such that*

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

Proof. By Theorem 4.7, $S_\Re(K) = \mathcal{R}(\Sigma_\Phi, \Phi)$ and the assumption that there is 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 $(f_{n_m}^2)_{m=1}^\infty$ 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 $\beta_{n_m} = 2^m \leq f_{n_m}^2$ for m such that $f_{n_m}^2 < \alpha_{n_m}$ and $\beta_n = \alpha_n$ for all other n. Then B is a Φ-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,

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

$$= \sum_{n \notin (n_m)_{m=1}^\infty} \frac{f_n^2}{\alpha_n} + \sum_{\substack{m \in \mathbb{N} \\ f_{n_m}^2 \geq \alpha_{n_m}}} \frac{f_{n_m}^2}{\alpha_{n_m}} + \sum_{\substack{m \in \mathbb{N} \\ f_{n_m}^2 < \alpha_{n_m}}} \frac{f_{n_m}^2}{2^m}$$

$$\geq \sum_{n \notin (n_m)_{m=1}^\infty} \frac{f_n^2}{\alpha_n} + \sum_{\substack{m \in \mathbb{N} \\ f_{n_m}^2 \geq \alpha_{n_m}}} 1 + \sum_{\substack{m \in \mathbb{N} \\ f_{n_m}^2 < \alpha_{n_m}}} 1$$

$$= \sum_{n \notin (n_m)_{m=1}^\infty} \frac{f_n^2}{\alpha_n} + \sum_{m=1}^\infty 1$$

$$= \infty,$$

which contradicts the assumption that $(f_n^2)_{n=1}^\infty$ is Σ_Φ-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 Φ-scaling of $H(K)$ and $K_{B,\Phi} \in \Sigma_\Phi$ because we have proved that $\sup_{n \geq 1} f_n^2 < \infty$. 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 Σ_Φ-approximately constant. $\qquad\square$

Recall from Section 3 that a scaled RKHS depends on the ordering of the orthonormal basis of $H(K)$. However, the sample support set of Theorem 4.8 does not depend on the ordering because the characterisation (4.4) is invariant to permutations. That is, let $\pi \colon \mathbb{N} \to \mathbb{N}$ be any permutation. Then $f = \sum_{n=1}^\infty f_n \phi_n = \sum_{n=1}^\infty f_{\pi(n)} \phi_{\pi(n)}$ 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 $(\phi_n)_{n=1}^\infty$ and $(\phi_{\pi(n)})_{n=1}^\infty$ 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*

$$a_n = \frac{a_n'}{(\sum_{l=n}^\infty a_l')^c}.$$

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 $0 \le c < 1$.

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*

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

Proof. For each $m \in \mathbb{N}$, define the sequence $(\bar{a}_{m,n})_{n=1}^\infty$ by setting $\bar{a}_{m,n} = \sum_{k=1}^m a_{k,n}$. Therefore $\bar{a}_{m,n} \le \bar{a}_{m+1,n}$ and $\sum_{n=1}^\infty \bar{a}_{m,n} = \sum_{k=1}^m \sum_{n=1}^\infty a_{k,n} < \infty$ for every m. Consequently, there is a strictly increasing sequence $(n_m)_{m=1}^\infty$ such that $\sum_{n=n_m}^\infty \bar{a}_{m,n} \le 2^{-m}$ for every $m \ge 2$. Set $a_n' = \bar{a}_{1,n}$ when $n < n_2$ and $a_n' = \bar{a}_{m,n}$ when $n_m \le n < n_{m+1}$ for $m \ge 2$. Then

$$\sum_{n=1}^\infty a_n' = \sum_{n=1}^{n_2-1} \bar{a}_{1,n} + \sum_{m=2}^\infty \sum_{n=n_m}^{n_{m+1}-1} \bar{a}_{m,n} \le \sum_{n=1}^\infty \bar{a}_{1,n} + \sum_{m=2}^\infty 2^{-m} < \infty$$

and, because $\bar{a}_{m,n} \le \bar{a}_{m+1,n}$ and $\bar{a}_{m,n} \ge a_{m,n}$, we have $a_n' \ge \bar{a}_{m,n} \ge a_{m,n}$ for all m and $n \ge n_m$. Finally, selecting $a_n = a_n'/(\sum_{l=n}^\infty a_l')^{1/2}$ yields, by Lemma 4.9, a convergent series such that

$$\lim_{n\to\infty} \frac{a_n}{a_{m,n}} \ge \lim_{n\to\infty} \frac{a_n}{a_n'} = \lim_{n\to\infty} \left(\sum_{l=n}^\infty a_n' \right)^{-1/2} = \infty$$

for every $m \in \mathbb{N}$. This proves the claim. $\qquad\square$

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

$$\sum_{n=1}^\infty \frac{1}{\alpha_{m,n}} < \infty \quad and \quad \sum_{n=1}^\infty \frac{1}{\beta_{m,n}} = \infty.$$

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

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

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, $H(K_{A',\Phi}) \subsetneq H(K_{A_m,\Phi})$ 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 Φ-scaling $A = (\alpha_n)_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$ and $H(K_{A,\Phi}) \subsetneq H(K_{A',\Phi})$. By Theorem 4.3, the set

$$F = H(K_{A,\Phi}) \setminus \bigcup_{m=1}^{\infty} H(K_{B_m,\Phi}) \subsetneq \bigcap_{m=1}^{\infty} H(K_{A_m,\Phi}) \setminus \bigcup_{m=1}^{\infty} H(K_{B_m,\Phi})$$

is of μ_X-measure one and thus $S_{\mathfrak{R}}(K) \subset F$ by definition. But Proposition 3.5 can be used again to construct a set $F' \subsetneq F$ such that $S_{\mathfrak{R}}(K) \subset F'$. Hence $S_{\mathfrak{R}}(K) \subsetneq F$. This completes the proof. $\square$

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 $\lim_{n \to \infty} \beta_{m,n}/\beta_{m+1,n} = 0$ for every $m \in \mathbb{N}$, then there exists $B = (\beta_n)_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} \beta_n^{-1} = \infty$ and $\lim_{n \to \infty} \beta_{m,n}/\beta_n = 0$ for every $m \in \mathbb{N}$. Although there is a Φ-scaling of $H(K)$ such that $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$, the divergence of $\sum_{n=1}^{\infty} \beta_n^{-1}$ does not have to imply that B is a Φ-scaling because *a divergent series can have terms arbitrarily larger than those of a convergent series* [19, p. 303]: there is a positive sequence $(\gamma_n)_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} \gamma_n^{-1} < \infty$ and $\liminf_{n \to \infty} \gamma_n/\beta_n = 0$. If $(\gamma_n)_{n=1}^{\infty}$ is a Φ-scaling of $H(K)$, we cannot directly deduce that so is B because $\beta_n \preceq \gamma_n$ fails.

4.4. On monotonely scaled RKHSs

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

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

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{m=0}^{\infty} (g_{m+1} - g_m) b_{g_m} \tag{4.5}$$

for any sequence $(b_n)_{n=1}^{\infty}$. 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 $\sum_{m=0}^{\infty} (g_{m+1} - g_m) b_{g_m}$ either both converge or diverge. That $(a_n)_{n=1}^{\infty}$ is Σ_M-approximately constant follows then from (4.5). The condition $g_{m+1} - g_m \leq C(g_m - g_{m-1})$, or equivalently $a_{g_m} \leq C a_{g_{m-1}}$, guarantees that a_{g_m} does not grow too fast in relation to g_m. The canonical example is obtained by setting $g_m = 2^m$ so that $a_{2^m} = 2^m$ and $C = 2$

and $\sum_{m=0}^{\infty}(g_{m+1} - g_m)b_{g_m} = \sum_{m=0}^{\infty} 2^m b_{2^m}$ converges or diverges with $\sum_{n=1}^{\infty} b_n$ by the standard Cauchy condensation test. Note that, because

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

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 Σ_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 $\sum_{n=1}^{\infty} n^{-\rho} < \infty$ if and only if $\rho > 1$, Theorem 4.3 implies that $\mathbb{P}[X \in H(K_{A_\rho, \Phi})] = 1$ if $\rho > 1$ and $\mathbb{P}[X \in H(K_{A_\rho, \Phi})] = 0$ if $\rho \leq 1$. Therefore the samples are "almost" contained in $H(K_{A_1, \Phi})$, while $H(K_{A_{1+\varepsilon}, \Phi})$ is a "small" RKHS which contains the samples for any "small" $\varepsilon > 0$. Thus the sample support set in Theorem 4.11 satisfies

$$S_{\mathfrak{R}}(K) \subsetneq \bigcap_{k=1}^{\infty} H(K_{A_{1+1/k}, \Phi}) \setminus H(K_{A_1, \Phi}),$$

where the set on the right-hand side is in $\mathcal{B}$ and has μ_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 μ_X-measure one. One such set is $F = H(K_{A, \Phi}) \setminus H(K_{A', \Phi})$ for the 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 $H(K_{A, \Phi}) \subset \bigcap_{k=1}^{\infty} H(K_{A_{1+1/k}, \Phi})$ and $H(K_{A_1, \Phi}) \subsetneq H(K_{A, \Phi})$.

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 $A_{\log(\rho)} = (\alpha_n)_{n=1}^{\infty}$ by $\alpha_n = (n+q)\log(n+q) \times \cdots \times \log_{p-1}(n+q)\log_p(n+q)^{\rho}$ and $A_{\log} = (\alpha'_n)_{n=1}^{\infty}$ by $\alpha'_n = (n+q)\log(n+q) \times \cdots \times \log_{p-1}(n+q)\log_p(n+q)$. It can be proved that $\sum_{n=1}^{\infty} \alpha_n^{-1} < \infty$ but $\sum_{n=1}^{\infty}(\alpha'_n)^{-1} = \infty$ [19, pp. 123, 280, 293]. Moreover,

$$\lim_{n \to \infty} \frac{n\log(n+1)^2}{(n+q)\log(n+q) \times \cdots \times \log_{p-1}(n+q)\log_p(n+q)^{\rho}} = \infty$$

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 μ_X-measure one which are smaller than the set F above. But Theorem 4.11 again demonstrates that for any set of μ_X-measure one constructed out of scalings $A_{\log(\rho)}$ and $A_{\log}$ 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 [17, Section 4.4] and [36] and 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]:

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

where $\phi_n(t) = \sqrt{2}(\pi n)^{-s/2}\sin(\pi nt)$. One can show that for even parameters the kernel is

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

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 $2s - 1$.
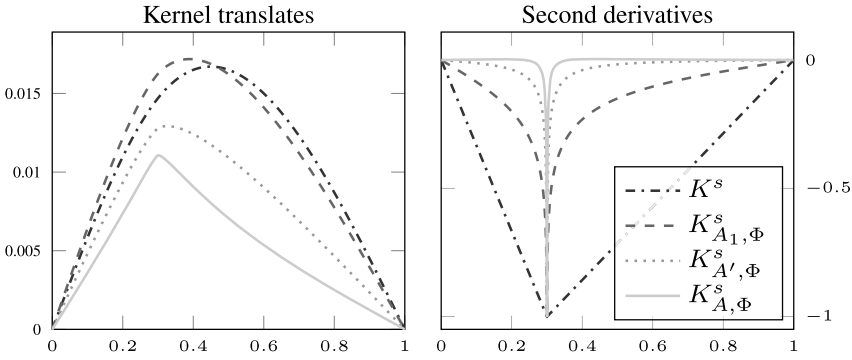
Figure 1. Translates (left) and their second derivatives (right) at $t' = 0.3$ of the kernel K^s and its three translates $K^s_{A_1,\Phi}$, $K^s_{A',\Phi}$ and $K^s_{A,\Phi}$ for $s = 4$. The second derivatives have been scaled so as to have -1 as their minimum. The scaled kernels were computed by truncating the series after 5,000 terms.

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

$$K^s_{A_\rho,\Phi}(t,t') = \frac{2}{\pi^s} \sum_{n=1}^{\infty} \frac{\sin(\pi nt)\sin(\pi nt')}{n^{s-\rho}} = \frac{1}{\pi^\rho} K^{s-\rho}(t,t'). \tag{5.2}$$

By Theorem 4.3, samples of $(X(t))_{t \in T} \sim \mathcal{GP}(0, K^s)$ 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 $2\int_0^1 \sin(\pi nt)\sin(\pi mt)\,dt = \delta_{nm}$ we see that the Mercer expansion with respect to the Lebesgue measure on $[0,1]$ of K^s is

$$K^s(t,t') = \sum_{n=1}^{\infty} \lambda_n \psi_n(t)\psi_n(t') \quad \text{with} \quad \lambda_n = \frac{1}{(\pi n)^s} \quad \text{and} \quad \psi_n(t) = \sqrt{2}\sin(\pi nt).$$

Therefore the θth power of $H(K^s)$, with $\theta = 1 - \rho/s$, equals $H(K^s_{A_\rho,\Phi})$ 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 $A = (\alpha_n)^\infty_{n=1}$ for $\alpha_n = n\log(n+1)^2$ and $A' = (\alpha'_n)^\infty_{n=1}$ for $\alpha'_n = n\log(n+1)$. These yield the scaled kernels

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

and

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

which do not appear to have closed form expressions. Because $\sum_{n=1}^{\infty} 1/(n\log(n+1)^2) < \infty$ but $\sum_{n=1}^{\infty} 1/(n\log(n+1)) = \infty$ the samples from $(X(t))_{t \in T} \sim \mathcal{GP}(0, K^s)$ are located in $H(K^s_{A,\Phi})$ but not in $H(K^s_{A',\Phi})$.

Translates at $t' = 0.3$ of the kernel K^s in (5.1) and the scaled kernels $K^s_{A_1,\Phi}$ in (5.2), $K^s_{A,\Phi}$ in (5.3) and $K^s_{A',\Phi}$ 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 $H(K^s_{A,\Phi})$ but not in $H(K^s_{A',\Phi})$ or $H(K^s_{A_1,\Phi})$. Moreover,

$$H(K^s) \subsetneq H(K^s_{A_1,\Phi}) \subsetneq H(K^s_{A',\Phi}) \subsetneq H(K^s_{A,\Phi}) \tag{5.5}$$

by Proposition 3.5. The second derivative of $K^s(\cdot, t')$ is Lipschitz while those of the scaled kernels are less well-behaved, though nevertheless continuous. How the second derivatives behave is indicative of the inclusions in (5.5): the larger the RKHS, the more severe the non-differentiability at $t = 0.3$.

5.2. Gaussian kernel

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

$$K(t,t') = \exp\left(-\frac{(t-t')^2}{2\ell^2}\right) = \exp\left(-\frac{t^2}{2\ell^2}\right)\exp\left(-\frac{(t')^2}{2\ell^2}\right)\sum_{n=0}^{\infty}\frac{1}{\ell^{2n}n!}(tt')^n \tag{5.6}$$

and the functions

$$\phi_n(t) = \frac{1}{\ell^n\sqrt{n!}}t^n\exp\left(-\frac{t^2}{2\ell^2}\right) \quad \text{for} \quad n \geq 0 \tag{5.7}$$

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

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

$$K_{A_\rho,\Phi}(t,t') = \exp\left(-\frac{t^2 + (t')^2}{2\ell^2}\right)\left(1 + \sum_{n=1}^{\infty}\frac{n^\rho}{\ell^{2n}n!}(tt')^n\right). \tag{5.8}$$

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

$$\begin{aligned}
K_{A_1,\Phi}(t,t') &= \exp\left(-\frac{t^2 + (t')^2}{2\ell^2}\right)\left(1 + \frac{tt'}{\ell^2}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{tt'}{\ell^2}\right)^n\right) \\
&= \exp\left(-\frac{t^2 + (t')^2}{2\ell^2}\right)\left(1 + \frac{tt'}{\ell^2}\exp\left(\frac{tt'}{\ell^2}\right)\right).
\end{aligned} \tag{5.9}$$

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 $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ with covariance kernel (5.6). For a few larger values of ρ we have

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

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

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

where $a = tt'/\ell^2$. 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 $\rho > 1$.
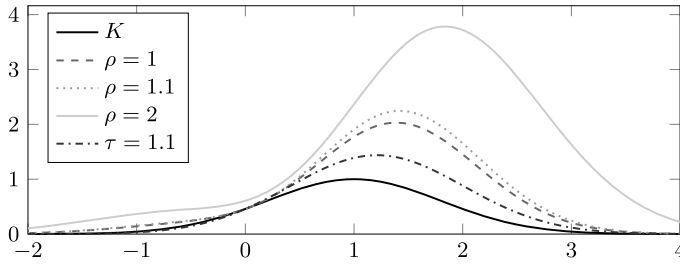
Figure 2. The translates at $t' = 1$ of the Gaussian kernel K in (5.6), the kernels $K_{A_\rho, \Phi}$ in (5.8) for $\rho \in \{1, 1.1, 2\}$, and the kernel in (5.10) with $\tau = 1.1$. The length-scale is $\ell = 0.8$.

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

$$K_{A,\Phi}(t, t') = \exp\left(-\frac{t^2 + (t')^2}{2\ell^2}\right) \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{\tau^2 t t'}{\ell^2}\right)^n = \exp\left(-\frac{t^2 + (t')^2}{2\ell^2} + \frac{\tau^2 t t'}{\ell^2}\right). \tag{5.10}$$

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 consruction 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

$$K_M^r(t, t') = \exp\left(-\frac{r^2(t^2 + (t')^2) - 2rtt'}{2(1 - r^2)}\right) = \sqrt{1 - r^2} \sum_{n=0}^{\infty} \frac{r^n}{n!} \mathrm{H}_n(t)\mathrm{H}_n(t'),$$

where H_n is the nth probabilists' Hermite polynomial. The RKHS of the Mehler kernel is analysed in [14].

A few of the kernels mentioned above are shown in Figure 2. By Proposition 3.5, their RKHSs satisfy

$$H(K) \subsetneq \underset{\rho = 1}{H(K_{A_\rho, \Phi})} \subsetneq \underset{\rho = 1.1}{H(K_{A_\rho, \Phi})} \subsetneq \underset{\rho = 2}{H(K_{A_\rho, \Phi})} \subsetneq \underset{A = (\tau^{2n})_{n=0}^{\infty}}{H(K_{A, \Phi})}$$

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

$$K(t, t') = \sum_{n=0}^{\infty} \frac{w_n}{(n!)^2}(tt')^n \tag{5.11}$$

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 $\mathbb{H}_1^2$ 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 $\phi_n(t) = t^n \sqrt{w_n}/n!$ 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 $(X(t))_{t \in T} \sim \mathcal{GP}(0, K)$ are in the RKHS of a different power series kernel

$$\bar{K}(t,t') = \sum_{n=0}^{\infty} \frac{\bar{w}_n}{(n!)^2} (tt')^n \quad \text{if and only if} \quad \sum_{n=0}^{\infty} \frac{w_n}{\bar{w}_n} < \infty.$$

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 $\sum_{n=0}^{\infty} w_n/\bar{w}_n = \sum_{n=0}^{\infty} n!/(n!)^2 < \infty$. Its samples are *not* in the RKHS of the kernel

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

where $w_0 = 1$ and $w_n = n!n$ for $n \geq 1$, because $\sum_{n=0}^{\infty} w_n/\bar{w}_n = 1 + \sum_{n=1}^{\infty} n^{-1} = \infty$. Note that this is essentially the same example as the one for the Gaussian kernel that involved the kernel (5.9).

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 [6,8]. In applications it is typical that the covariance kernel has parameters, such as the length-scale parameter $\ell > 0$ 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, [31, Section 5.4.1] or [33, Chapter 3]. This section uses Theorem 4.3 to explain the behaviour of the maximum likelihood estimate of the kernel scaling parameter that Xu and Stein [41] observed recently.

Let $K \colon T \times T \to \mathbb{R}$ be a positive-definite kernel and $K_\sigma = \sigma^2 K$ a kernel parametrised by a non-negative scaling parameter σ. Suppose that the data consist of evaluations of a function $f \colon T \to \mathbb{R}$ 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 easy to compute that the maximum likelihood estimate $\hat{\sigma}_{f,N}$ of σ is (see the references above)

$$\hat{\sigma}_{f,N} = \sqrt{\frac{\mathsf{f}_N^\mathsf{T} \mathsf{K}_N^{-1} \mathsf{f}_N}{N}},$$

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

$$\mathbb{E}_f \big[\hat{\sigma}_{f,N}^2 \big] = \frac{\mathbb{E}_f \big[\mathsf{f}_N^\mathsf{T} \mathsf{K}_N^{-1} \mathsf{f}_N \big]}{N} = \frac{\mathbb{E}_f \big[\mathrm{tr}(\mathsf{f}_N \mathsf{f}_N^\mathsf{T} \mathsf{K}_N^{-1}) \big]}{N} = \frac{\mathrm{tr}(\sigma_0^2 \mathsf{K}_N \mathsf{K}_N^{-1})}{N} = \sigma_0^2.$$

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 $f(t) = \exp(-t^2/(2\ell^2)) \sum_{n=0}^\infty t^n/(\ell^n \sqrt{n!})$ 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 from 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,

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

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

$$\lim_{N \to \infty} \hat{\sigma}_{f,N} = 0 \ \text{ if } \ f \in H(K_0) \quad \text{and} \quad \lim_{N \to \infty} \hat{\sigma}_{f,N} = \infty \ \text{ if } \ f \in H(K_1) \setminus H(K_2) \tag{6.1}$$

under some assumptions on T and the points t_i. We now show that the behaviour of $\hat{\sigma}_{f,N}$ 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 $C_{\ell,1}$,

$$\hat{\sigma}_{f,N}^2 \sim C_{\ell,0} N^{-1/2} \ \text{ if } \ f \equiv 1 \quad \text{and} \quad \liminf_{N \to \infty} N^{-1/2} \hat{\sigma}_{f,N}^2 \geq C_{\ell,1} \ \text{ if } \ f(t) = t$$

as $N \to \infty$. Furthermore, they conjectured that

$$\hat{\sigma}_{f,N}^2 \sim C_{\ell,p} N^{p-1/2} \quad \text{as} \quad N \to \infty \tag{6.2}$$

if $f(t) = t^p$ for any $p \geq 0$ and a positive constant $C_{\ell,p}$.[4] 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]

---

[4] As pointed out by Dette and Zhigljavsky [7, Section 4.1], it appears that the constant $C_{\ell,p} = \ell^{2p}/(\sqrt{2\pi}(p + 1/2))$ 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 $C_{\ell,p} = 2^p \ell^{2p}/(\sqrt{\pi}(p + 1/2))$.

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) as

$$t^p = \exp\left(\frac{t^2}{2\ell^2}\right) t^p \exp\left(-\frac{t^2}{2\ell^2}\right) = \sum_{n=0}^{\infty} \frac{t^{2n+p}}{2^n \ell^{2n} n!} \exp\left(-\frac{t^2}{2\ell^2}\right) = \ell^p \sum_{n=0}^{\infty} \frac{\sqrt{(2n+p)!}}{2^n n!} \phi_{2n+p}(t). \qquad (6.3)$$

Let $A = (\alpha_n)_{n=1}^{\infty}$ be a Φ-scaling of $H(K)$. If the pth monomial is to be an element of the scaled RKHS $H(K_{A,\Phi})$ it must, by Proposition 3.2, hold that

$$\sum_{n=0}^{\infty} \alpha_{2n+p}^{-1} \frac{(2n+p)!}{2^{2n}(n!)^2} < \infty. \qquad (6.4)$$

From Stirling's formula and $\lim_{n\to\infty}(1 + \frac{p}{2n})^{2n} = e^p$ we get

$$\frac{(2n+p)!}{2^{2n}(n!)^2} \sim \frac{\sqrt{2\pi}(2n+p)^{2n+p+1/2} e^{-(2n+p)}}{2\pi 2^{2n} n^{2n+1} e^{-2n}} = \frac{1}{\sqrt{2\pi} e^p} \frac{(2n+p)^{2n+p+1/2}}{(2n)^{2n} n}$$

$$= \frac{1}{\sqrt{2\pi} e^p} \left(1 + \frac{p}{2n}\right)^{2n} \frac{(2n+p)^{p+1/2}}{n}$$

$$\sim \frac{2^p}{\sqrt{\pi}} n^{p-1/2}$$

as $n \to \infty$. Therefore (6.4) holds if and only if $\sum_{n=1}^{\infty} \alpha_{2n+p}^{-1} n^{p-1/2} < \infty$. If we constraint ourselves to hyperharmonic scalings this is clearly equivalent to $\alpha_n = n^\rho$ for $\rho > p + 1/2$. But since $\sum_{n=1}^{\infty} n^{-\rho} < \infty$ 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 $A = (n^\rho)_{n=0}^{\infty}$ and any $\rho \in (1/2, 1]$, f 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 $A = (n^\rho)_{n=0}^{\infty}$ and $A' = (n^\sigma)_{n=0}^{\infty}$, 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 $f \in H(K_1) \setminus H(K_2)$.

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

Funding

References

[1] Adler, R.J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **12**. Hayward, CA: IMS. MR1088478

[2] Bogachev, V.I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Providence, RI: Amer. Math. Soc. MR1642391 https://doi.org/10.1090/surv/062

[3] Bull, A.D. (2011). Convergence rates of efficient global optimization algorithms. *J. Mach. Learn. Res.* **12** 2879–2904. MR2854351

[4] Cavoretto, R., Fasshauer, G.E. and McCourt, M. (2015). An introduction to the Hilbert-Schmidt SVD using iterated Brownian bridge kernels. *Numer. Algorithms* **68** 393–422. MR3304852 https://doi.org/10.1007/s11075-014-9850-z

[5] Cialenco, I., Fasshauer, G.E. and Ye, Q. (2012). Approximation of stochastic partial differential equations by a kernel-based collocation method. *Int. J. Comput. Math.* **89** 2543–2561. MR3004665 https://doi.org/10.1080/00207160.2012.688111

[6] Cockayne, J., Oates, C.J., Sullivan, T.J. and Girolami, M. (2019). Bayesian probabilistic numerical methods. *SIAM Rev.* **61** 756–789. MR4027836 https://doi.org/10.1137/17M1139357

[7] Dette, H. and Zhigljavsky, A.A. (2021). Reproducing kernel Hilbert spaces, polynomials, and the classical moment problem. *SIAM/ASA J. Uncertain. Quantificat.* **9** 1589–1614. MR4334548 https://doi.org/10.1137/21M1394965

[8] Diaconis, P. (1988). Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)* 163–175. New York: Springer. MR0927099

[9] Driscoll, M.F. (1973). The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrsch. Verw. Gebiete* **26** 309–316. MR0370723 https://doi.org/10.1007/BF00534894

[10] Engl, H.W., Hanke, M. and Neubauer, A. (1996). *Regularization of Inverse Problems. Mathematics and Its Applications* **375**. Dordrecht: Kluwer Academic. MR1408680

[11] Flaxman, S., Sejdinovic, D., Cunningham, J. and Filippi, S. (2016). Bayesian learning of kernel embeddings. In *Uncertainty in Artificial Intelligence* 182–191.

[12] Fortet, R. (1973). Espaces à noyau reproduisant et lois de probabilités des fonctions aléatoires. *Ann. Inst. Henri Poincaré B, Calc. Probab. Stat.* **9** 41–58. MR0345193

[13] Gualtierotti, A.F. (2015). *Detection of Random Signals in Dependent Gaussian Noise*. Cham: Springer. MR3445605 https://doi.org/10.1007/978-3-319-22315-5

[14] Irrgeher, C. and Leobacher, G. (2015). High-dimensional integration on $\mathbb{R}^d$, weighted Hermite spaces, and orthogonal transforms. *J. Complexity* **31** 174–205. MR3305992 https://doi.org/10.1016/j.jco.2014.09.002

[15] Kallianpur, G. (1970). Zero-one laws for Gaussian processes. *Trans. Amer. Math. Soc.* **149** 199–211. MR0266293 https://doi.org/10.2307/1995672

[16] Kallianpur, G. (1971). Abstract Wiener processes and their reproducing kernel Hilbert spaces. *Z. Wahrsch. Verw. Gebiete* **17** 113–123. MR0281242 https://doi.org/10.1007/BF00538863

[17] Kanagawa, M., Hennig, P., Sejdinovic, D. and Sriperumbudur, B.K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. Available at arXiv:1807.02582v1.

[18] Karvonen, T., Wynne, G., Tronarp, F., Oates, C. and Särkkä, S. (2020). Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. *SIAM/ASA J. Uncertain. Quantificat.* **8** 926–958. MR4130422 https://doi.org/10.1137/20M1315968

[19] Knopp, K. (1951). *Theory and Application of Infinite Series*, 2nd ed. Blackie & Son.

[20] Kreĭn, S.G. and Petunin, J.I. (1966). Scales of Banach spaces. *Russian Math. Surveys* **21** 85–159. MR0193499

[21] Le Page, R. (1973). Subgroups of paths and reproducing kernels. *Ann. Probab.* **1** 345–347. MR0350835 https://doi.org/10.1214/aop/1176996990

[22] Loh, W.-L. and Lam, T.-K. (2000). Estimating structured correlation matrices in smooth Gaussian random field models. *Ann. Statist.* **28** 880–904. MR1792792 https://doi.org/10.1214/aos/1015952003

[23] Lukić, M.N. (2004). Integrated Gaussian processes and their reproducing kernel Hilbert spaces. In *Stochastic Processes and Functional Analysis*. *Lecture Notes in Pure and Applied Mathematics* **238** 241–263. New York: Dekker. MR2059910

[24] Lukić, M.N. and Beder, J.H. (2001). Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Trans. Amer. Math. Soc.* **353** 3945–3969. MR1837215 https://doi.org/10.1090/S0002-9947-01-02852-5

[25] Minh, H.Q. (2010). Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory. *Constr. Approx.* **32** 307–338. MR2677883 https://doi.org/10.1007/s00365-009-9080-0

[26] Mörters, P. and Peres, Y. (2010). *Brownian Motion. Cambridge Series in Statistical and Probabilistic Mathematics* **30**. Cambridge: Cambridge Univ. Press. MR2604525 https://doi.org/10.1017/CBO9780511750489

[27] Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I Technical Report 23, Applied Mathematics and Statistics Laboratory, Stanford Univ.

[28] Parzen, E. (1963). Probability density functionals and reproducing kernel Hilbert spaces. In *Proc. Sympos. Time Series Analysis (Brown Univ., 1962)* 155–169. New York: Wiley. MR0149634

[29] Paulsen, V.I. and Raghupathi, M. (2016). *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*. *Cambridge Studies in Advanced Mathematics* **152**. Cambridge: Cambridge Univ. Press. MR3526117 https://doi.org/10.1017/CBO9781316219232

[30] Rakotomamonjy, A. and Canu, S. (2005). Frames, reproducing kernels, regularization and learning. *J. Mach. Learn. Res.* **6** 1485–1515. MR2249863

[31] Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press. MR2514435

[32] Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. MR1041765

[33] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*. *Springer Series in Statistics*. New York: Springer. MR2160708 https://doi.org/10.1007/978-1-4757-3799-8

[34] Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Process. Appl.* **120** 1879–1897. MR2673978 https://doi.org/10.1016/j.spa.2010.05.009

[35] Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. *Springer Series in Statistics*. New York: Springer. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

[36] Steinwart, I. (2019). Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *Potential Anal.* **51** 361–395. MR4023468 https://doi.org/10.1007/s11118-018-9715-5

[37] Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer.

[38] Steinwart, I., Hush, D. and Scovel, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. Inf. Theory* **52** 4635–4643. MR2300845 https://doi.org/10.1109/TIT.2006.881713

[39] Steinwart, I. and Scovel, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. MR2914365 https://doi.org/10.1007/s00365-012-9153-3

[40] van der Vaart, A. and van Zanten, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028

[41] Xu, W. and Stein, M.L. (2017). Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA J. Uncertain. Quantificat.* **5** 138–175. MR3601677 https://doi.org/10.1137/15M105358X

[42] Xu, Y. and Zhang, H. (2009). Refinement of reproducing kernels. *J. Mach. Learn. Res.* **10** 107–140. MR2476735

[43] Zhang, H. and Zhao, L. (2013). On the inclusion relation of reproducing kernel Hilbert spaces. *Anal. Appl. (Singap.)* **11** 1350014. MR3037130 https://doi.org/10.1142/S0219530513500140

[44] Zwicknagl, B. (2009). Power series kernels. *Constr. Approx.* **29** 61–84. MR2465291 https://doi.org/10.1007/s00365-008-9012-4

[45] Zwicknagl, B. and Schaback, R. (2013). Interpolation and approximation in Taylor spaces. *J. Approx. Theory*
171 65–83. MR3053717 https://doi.org/10.1016/j.jat.2013.03.006