Appendix I

Gaussian Processes

This appendix gives background on Gaussian processes and Gaussian random elements in Banach spaces. The focus is on properties that are important for their use as priors, as in Chapter 11.

For basic definitions and examples of Gaussian processes see Section 11.1.

I.1 Sample Path Properties

Sample path properties of a Gaussian process $W = (W_t : t \in T)$, such as boundedness, continuity or differentiability, are (of course) determined by the covariance kernel $K(s,t) = \operatorname{cov}(W_s, W_t)$ of the process. In this section we discuss continuity for an abstract index set T, and next differentiability for processes indexed by Euclidean space.

Although a precise characterization of continuity and boundedness in terms of majorizing measures is available (see e.g. Talagrand 1987, 1992), the following result in terms of the entropy of the index set with respect to the *intrinsic semimetric* ρ is a good compromise between generality and simplicity. The square of this metric is

$$\rho^2(s,t) = \operatorname{var}(W_s - W_t) = K(t,t) + K(s,s) - 2K(s,t).$$

Proposition I.1 (Continuity) *If $(W_t : t \in T)$ is a separable, mean-zero Gaussian process with intrinsic metric ρ, then for any $\delta > 0$,*

$$\mathrm{E}\Big[\sup_{\rho(s,t)\le\delta} |W_s - W_t|\Big] \lesssim \int_0^\delta \sqrt{\log D(\epsilon, T, \rho)}\, d\epsilon. \tag{I.1}$$

Furthermore, for $J(\delta)$ the integral on the right side,

$$\mathrm{E} \sup_{\rho(s,t)<\delta} \frac{|W_s - W_t|}{J(\rho(s,t))} < \infty.$$

Consequently, $|W_s - W_t| = O(J(\rho(s,t)))$, uniformly in (s,t) with $\rho(s,t) \to 0$, almost surely.

Proof The display is known as *Dudley's bound*, motivated from Dudley (1967); see for instance van der Vaart and Wellner (1996), Corollary 2.2.8 for a proof. The second inequality is equivalent to the finiteness of the supremum of the absolute values of the Gaussian process $(s,t) \mapsto (W_s - W_t)/J(\rho(s,t))$, which was also established in Dudley (1967), and is stronger than (I.1). ☐

Proposition I.2 (Modulus) *If $(W_t : t \in T)$ is a mean-zero Gaussian process indexed by a compact set $T \subset \mathbb{R}^d$ with $\mathrm{E}\|W_s - W_t\|^2 \leq \|s - t\|^{2\alpha}$, for all $s, t \in T$ and some $\alpha \in (0, 1]$, then W possesses a version with continuous sample paths such that $|W_s - W_t| = O(\|s - t\|^\alpha \log_- \|s - t\|)$, uniformly in (s, t) with $\|s - t\| \to 0$, almost surely.*

Proof Because the intrinsic metric ρ of W is by assumption bounded above by the Euclidean metric to the power α, the entropy of a compact interval is bounded above by a multiple of $\epsilon^{-d/\alpha}$. Hence the entropy integral is of the order $\int_0^\delta \sqrt{\log_- \epsilon}\, d\epsilon \sim \delta\sqrt{\log_- \delta}$, as $\delta \to 0$. As the process is continuous in probability, there exists a separable version of the process, by general results on stochastic processes. The result follows by applying the preceding proposition to such a version. ☐

For a multi-index $j = (j_1, \ldots, j_d)$ of nonnegative integers , let D^j denote the mixed partial derivative operator $\partial^{|j|}/\partial t_1^{j_1} \cdots \partial t_d^{j_d}$. Furthermore, for a function K of two arguments, let $D_s^i D_t^j K(s, t)$ denote the function obtained by differentiating the function i times with respect to s and j times with respect to t.

Proposition I.3 (Differentiability) *If the partial derivatives $(s, t) \mapsto D_s^j D_t^j K(s, t)$ of order j of the covariance kernel K of the mean-zero Gaussian process $(W_t : t \in T)$, for T an interval in $\mathbb{R}^d$, exist and are Lipschitz continuous of order $\alpha > 0$, then W possesses a version whose sample paths are partially differentiable up to order j with jth order derivative that is Lipschitz of order a, for any $a < \alpha$. Furthermore, the derivative process $D^j W$ is Gaussian with covariance function $\mathrm{cov}(D^j W_s, D^j W_t) = D_s^j D_t^j K(s, t)$.*

Proof First consider the case that $d = 1$ and $j = 1$ and set $Y_h^t = W_{t+h} - W_t$, for given t and h. By linearity of the covariance and continuous differentiability of K, for every s, t, g, h,

$$\begin{aligned}\mathrm{cov}(Y_g^s, Y_h^t) &= K(s + g, t + h) - K(s + g, t) - K(s, t + h) + K(s, t)\\ &= \int_s^{s+g} \int_t^{t+h} D_u D_v K(u, v)\, dv\, du.\end{aligned}$$

Then $\mathrm{var}(Y_h^t/h - Y_{h'}^t/h') = \mathrm{var}(Y_h^t/h) + \mathrm{var}(Y_{h'}^t/h') - 2\,\mathrm{cov}(Y_h^t/h, Y_{h'}^t/h') \to 0$ as $h, h' \to 0$, as the two variances and the covariance on the right all tend to $D_s D_t K(s, t)_{|s=t}$, by continuity of $D_s D_t K$. Thus Y_h^t/h is a Cauchy net and converges in $\mathbb{L}_2$ to a limit; denote this by $\dot W_t$. The latter process is mean-zero Gaussian with covariance function $\mathrm{cov}(\dot W_s, \dot W_t) = \lim \mathrm{cov}(Y_h^s/h, Y_h^t/h) = D_s D_t K(s, t)$, by the preceding display. By assumption the latter function is Lipschitz of order α (or of order 1 if $j > 1$), and hence the process $\dot W_t$ possesses a version that is Lipschitz of order $a < \alpha$, by Proposition I.2. We claim that this version $\dot W_t$ is also a pathwise derivative of $t \mapsto W_t$. Indeed, for arbitrary $t_0 \in T$ the process $V_t := W_{t_0} + \int_{t_0}^t \dot W_s\, ds$ is pathwise continuously differentiable, and can be seen to be also differentiable in $\mathbb{L}_2$, with derivative $\dot W_t$. Thus the process $W_t - V_t$ possesses $\mathbb{L}_2$-derivative equal to 0 and vanishes at t_0. This implies that the function $t \mapsto \mathrm{E}[(V_t - W_t)H]$ is continuously differentiable, for any square-integrable H, with derivative $\mathrm{E}[0H] = 0$, and hence is zero by the mean value theorem. Thus $V_t = W_t$ almost surely.

We can proceed to $j > 1$ by induction.

For $d > 1$ we can apply the argument coordinatewise, keeping the other coordinates fixed. This shows that the process is partially differentiable with continuous partial derivatives. Hence it is totally differentiable. □

Proposition I.4 (Stationary processes) *Suppose that $(W_t : t \in \mathbb{R}^d)$ is a mean-zero stationary Gaussian process with spectral measure μ.*

(i) *If $\int \|\lambda\|^{2\alpha d}\, d\mu(\lambda) < \infty$, then W has a version whose sample paths are partially differentiable up to order the biggest integer k strictly smaller than α with partial derivatives of order k that are Lipschitz of order $a - k$, for any $a < \alpha$.*
(ii) *If $\int e^{c\|\lambda\|}\, d\mu(\lambda) < \infty$ for some $c > 0$, then W has a version with analytic sample paths.*

Proof By the dominated convergence theorem, for any multi-indices j and h with $\sum_l (j_l + h_l) \leq 2k$,

$$D_s^j D_t^h K(s,t) = \int (i\lambda)^j (-i\lambda)^h e^{i\langle (s-t),\lambda\rangle}\, d\mu(\lambda).$$

Furthermore, the right-hand side is Lipschitz in (s,t) of order $\alpha - k$. Thus (i) follows from Proposition I.3.

For the proof of (ii) we note that the function $K(s,t) = \int e^{i\langle s-\bar{t},\lambda\rangle}\, d\mu(\lambda)$ is now also well defined for complex-valued s, t with absolute imaginary parts smaller than $c/2$ (and $\bar{t}$ the complex conjugate of t). The function is conjugate-symmetric and nonnegative-definite and hence defines a covariance function of a (complex-valued) stochastic process $(W_t : t \in T)$, indexed by $T = \{t \in \mathbb{C} : |\operatorname{Im} t| < c/2\}$. By an extension of Proposition I.3 this can be seen to have sample paths with continuous partial derivatives, which satisfy the Cauchy-Riemann equations, and hence the process is differentiable on its complex domain. □

I.2 Processes and Random Elements

If the sample paths of a Gaussian stochastic process $W = (W_t : t \in T)$ belong to a Banach space $\mathbb{B}$ of functions $z : T \to \mathbb{R}$, then W can be viewed as a map from the underlying probability space into $\mathbb{B}$. If this map is Borel measurable, then W induces a distribution on the Borel σ-field of $\mathbb{B}$. In Definition 11.2 W is defined to be a *Gaussian random element* if $b^*(W)$ is normally distributed, for every element b^* of the dual space $\mathbb{B}^*$. In this section we link the two definitions of being Gaussian, as a process and as a Borel measurable map.

We start with considering Borel measurability of a stochastic process.

Lemma I.5 *If the sample paths $t \mapsto W_t$ of the stochastic process $W = (W_t : t \in T)$ belong to a separable subset of a Banach space and the norm $\|W - w\|$ is a random variable for every w in the subset, then W is a Borel measurable map in this space.*

Proof The condition implies that $\{W \in B\}$ is measurable for every open or closed ball B in the given subset $\mathbb{B}_0$ in which W takes its values. Thus W is measurable in the σ-field generated in $\mathbb{B}_0$ by these balls. As $\mathbb{B}_0$ is separable by assumption, its Borel σ-field is equal to the ball σ-field (see e.g. Chapter 1.7 in van der Vaart and Wellner 1996). Thus W is a

Borel measurable map in $\mathbb{B}_0$. As the trace of the Borel σ-field in the encompassing Banach space is the Borel σ-field in $\mathbb{B}_0$, this is then also true for W as a map in $\mathbb{B}$. $\square$

The lemma applies for instance to the space of continuous functions on a compact metric space, where the supremum norm can be seen to be measurable by reducing the supremum to a countable, dense set. The assumption of the lemma that the sample paths belong to a *separable* subset is not harmless. For instance, the Hölder norm of a stochastic process with sample paths in a Hölder space $\mathfrak{C}^\alpha[0, 1]$ is easily seen to be measurable, but the Hölder space itself is not separable relative to its norm. The sample paths must be contained in a smaller set for the lemma to apply; for instance, a Hölder space of higher order $\beta > \alpha$ would do.

If a Gaussian stochastic process is a Borel measurable map in a Banach function space, then it is still not clear that it also fulfills the requirement of a Gaussian random element that every variable $b^*(W)$ is Gaussian. The following lemma addresses this for the Banach space of bounded functions.

Lemma I.6 *If the stochastic process $W = (W_t : t \in T)$ is a Borel measurable random element in a separable subset of the Banach space $\mathfrak{L}_\infty(T)$ equipped with the supremum norm, then W is a Gaussian random element in this space.*

Proof Every coordinate projection π_t, defined by $\pi_t z = z(t)$, is an element of the dual space $\mathfrak{L}_\infty(T)^*$, and so are linear combinations of coordinate projections. The assumption that W is Gaussian implies that $b^*(W)$ is Gaussian for every such linear combination $b^* = \sum_i \alpha_i \pi_{s_i}$. We shall show that a general b^* is a pointwise limit of linear combinations of coordinate projections, at least on a set in which W takes its values. Then Gaussianity of $b^*(W)$ follows from the fact an almost sure limit of Gaussian variables is Gaussian.

We may assume without loss of generality that the separable subset of $\mathfrak{L}_\infty(T)$ in which W takes its values is complete. Then W is automatically a tight random element, and it is known that there exists a semimetric ρ on T under which T is totally bounded and such that W takes its values in the subspace $\mathfrak{UC}(T, \rho)$ of functions $f : T \to \mathbb{R}$ that are uniformly continuous relative to ρ (e.g. van der Vaart and Wellner 1996, Lemma 1.5.9). Thus we may assume without loss of generality that W takes its values in $\mathfrak{UC}(T, \rho)$ for such a semimetric ρ.

Only the restriction to $\mathfrak{UC}(T, \rho)$ of a given element of $\mathfrak{L}_\infty(T)^*$ is now relevant, and this is contained in $\mathfrak{UC}(T, \rho)^*$. By the Riesz representation theorem an arbitrary element of $\mathfrak{UC}(T, \rho)^*$ is a map $f \mapsto \int \bar{f}(t)\, d\mu(t)$ for a signed Borel measure μ on the completion $\bar{T}$ of T and $\bar{f} : \bar{T} \to \mathbb{R}$ the continuous extension of f. Because T is totally bounded we can write it for each $m \in \mathbb{N}$ as a finite union of sets of diameter smaller than $1/m$. If we define μ_m as the measure obtained by concentrating the masses of μ on the partitioning sets in a fixed, single point in the partitioning set, then $\int \bar{f}\, d\mu_m \to \int \bar{f}\, d\mu$ as $m \to \infty$, for each $f \in \mathfrak{UC}(T, \rho)$. The map $f \mapsto \int \bar{f}\, d\mu_m$ is a linear combination of coordinate projections. It follows that for any $b^* \in \mathfrak{UC}(T, \rho)^*$ there exists a sequence b_m^* of linear combinations of coordinate projections that converges pointwise on $\mathfrak{UC}(T, \rho)$ to b^*. $\square$

The preceding proof is based on approximating an arbitrary element of the dual space $\mathbb{B}^*$ pointwise by a sequence of linear combinations of coordinate projections. Thus a sufficient general condition is that the linear span of the coordinate projections $z \mapsto z(t)$ form a weak-* dense subset of the dual space. However, as the proof shows weak-* approximation need only hold on a subset of $\mathbb{B}$ where W takes its values.

Consider the special example of the Hölder space $\mathfrak{C}^\alpha([0,1]^d)$. A simple sufficient condition is that the sample paths of the process are smoother than α.

Lemma I.7 *A Gaussian process $W = (W_t : t \in [0,1]^d)$ with sample paths in $\mathfrak{C}^\beta([0,1]^d)$ is a Gaussian random element in $\mathfrak{C}^\alpha([0,1]^d)$, for $\alpha < \beta$.*

Proof For $\alpha < \beta$ the space $\mathfrak{C}^\beta([0,1]^d)$ is a separable subset of $\mathfrak{C}^\alpha([0,1]^d)$ for the $\mathfrak{C}^\alpha$-norm, and the norm $\|W - w\|_{\mathfrak{C}^\alpha}$ can be seen to be a random variable for every $w \in \mathfrak{C}^\beta([0,1]^d)$ by writing it as a countable supremum. Therefore W is Borel measurable by Lemma I.5. To show that it is Gaussian, for simplicity restrict to the subspace $\mathfrak{C}_0^\alpha[0,1]^d$ of functions that vanish at 0. A function f in this space can be identified with the function $v_f : [0,1]^{2d} \setminus D \to \mathbb{R}$, for $D = \{(t,t) : t \in [0,1]^d\}$, defined by

$$v_f(s,t) = \frac{f(s) - f(t)}{|s-t|^\alpha}.$$

In fact this identification gives an isometry of $\mathfrak{C}_0^\alpha([0,1]^d)$ onto the set V of all functions v_f equipped with the uniform norm. Since every sample of W is in $\mathfrak{C}^\beta([0,1]^d)$, the image v_W takes its values in the subspace $V_0 \subset V$ of functions that can be continuously extended to a function in $\mathfrak{C}([0,1]^{2d})$ that vanishes on D. The restriction of $b^* \in V^*$ to V_0 is a continuous linear map on V_0 that can be extended to an element of the dual space of $\mathfrak{C}([0,1]^{2d})$, by the Hahn-Banach theorem. By the Riesz representation theorem it is representable by a finite Borel measure μ on $[0,1]^{2d}$. Putting it all together we see that $b^*(W) = \int (W(s) - W(t))/|s-t|^\alpha \, d\mu(s,t)$. We now approximate μ by a sequence of discrete measures and finish as in the proof of Lemma I.6. ∎

I.3 Probability Bounds

The distribution of a Gaussian process or random element is tightly concentrated around its mean. *Borell's inequality*, Proposition 11.17, is an expression of this fact. The following inequality is often also called Borell's inequality, or *Borell-Sudakov-Tsirelson inequality*, but concerns the concentration of the norm of the process. The inequality applies both to random elements in a Banach space and to stochastic processes.

For W a mean-zero Gaussian element in a separable Banach space, let $M(W)$ be the median of $\|W\|$, and

$$\sigma^2(W) = \sup_{b^* \in \mathbb{B}^* : \|b^*\| = 1} \mathrm{E}\big[b^*(W)^2\big]. \tag{I.2}$$

Alternatively, for $W = (W_t : t \in T)$ a separable mean-zero Gaussian process, assume that $\|W\| := \sup_t |W_t|$ is finite almost surely, and let $M(W)$ be the median of $\|W\|$, and

$$\sigma^2(W) = \sup_t \mathrm{var}(W_t).$$

Proposition I.8 (Borell-Sudakov-Tsirelson) *Let W be a mean-zero Gaussian random element in a Banach space or a mean-zero separable Gaussian process such that $\|W\| < \infty$ almost surely. Then $\sigma(W) \leq M(W)/\Phi^{-1}(3/4)$, and for any $x > 0$,*

$$\mathrm{P}\Big(\big|\|W\| - M(W)\big| > x\Big) \leq 2\big[1 - \Phi(x/\sigma(W))\big],$$
$$\mathrm{P}\Big(\big|\|W\| - \mathrm{E}\|W\|\big| > x\Big) \leq 2e^{-x^2/(2\sigma^2(W))}.$$

Proof For W a Gaussian process, see for instance van der Vaart and Wellner (1996), Proposition A.2.1. The upper bound on $\sigma(W)$ follows from the facts that $X_t/\operatorname{sd}(X_t) \sim \operatorname{Nor}(0, 1)$ and $\mathrm{P}(|X_t| \leq M(X)) \geq 1/2$, for every t. The Banach space version can be reduced to the stochastic process version of the process $(b^*(W)\colon \|b^*\| = 1)$. The bound on the upper tail $\mathrm{P}(\|W\| - M(W) > x)$ is also a consequence of Borell's inequality, Proposition 11.17, upon taking $\epsilon = M(W)$ and $M = x/\sigma(W)$. □

The remarkable fact is that these bounds are independent of the size or complexity of the index set T or the Banach space. For any index set the tails of the distribution of $\|W\|$ away from its median or mean are smaller than Gaussian. On the other hand, the size of this median or mean, and hence the location of the distribution of $\|W\|$, strongly depends on the complexity of T or the Banach space. Proposition I.1 gives an upper bound in terms of the metric entropy of T.

Somewhat remarkable too is that the variance of the Gaussian tail bound is equal to $\sigma^2(W)$, which is the maximal variance of a variable W_t or $b^*(W)$. Thus the bound is no worse than the Gaussian tail bound for a single (worse case) variable in the supremum. By integrating $2x$ times the second bound we see that $\operatorname{var}[\|W\|] \leq 4\sigma^2(W)$, so that the variance of the norm is bounded by a multiple of the maximal variance.

By integrating the first inequality it is seen that $|\mathrm{E}\|W\| - M(W)| \leq \sigma(W)\sqrt{2/\pi}$, which shows that the mean is finite as soon as $\|W\|$ is a finite random variable. By integrating px^{p-1} times the first inequality we see that all moments of $\|W\|$ are finite, and can be bounded in terms of the first two moments, in the same way as for a univariate Gaussian variable.

I.4 Reproducing Kernel Hilbert Space

In Section 11.2 the *reproducing kernel Hilbert space* (RKHS) $\mathbb{H}$ attached to a Gaussian variable is defined both for Gaussian processes and for Gaussian random elements in a separable Banach space. In this section we give further background.

The definition of the Banach space RKHS uses the *Pettis integral* of a random element X with values in a Banach space $\mathbb{B}$. This is defined as the unique element $\mu \in \mathbb{B}$ such that $b^*(\mu) = \mathrm{E}[b^*(X)]$ for every $b^* \in \mathbb{B}^*$; it is denoted by $\mu = \mathrm{E}(X)$. The following lemma gives a condition for existence of this expectation.

Lemma I.9 *If X is a Borel measurable map in a separable Banach space $\mathbb{B}$ with $\mathrm{E}\|X\| < \infty$, then there exists an element $\mu \in \mathbb{B}$ such that $b^*(\mu) = \mathrm{E}[b^*(X)]$, for every $b^* \in \mathbb{B}^*$.*

Proof As a random element X taking values in a complete, separable metric space is automatically tight (cf. Parthasarathy 2005, or van der Vaart and Wellner 1996, 1.3.2), for any $n \in \mathbb{N}$ there exists a compact set K such that $\mathrm{E}[\|X\| \mathbb{1}_{\{X \notin K\}}] < n^{-1}$. Partition K in finitely many sets $B_1, \ldots, B_k$ of diameter less than n^{-1}. For increasing n, these partitions can be chosen successive refinements without loss of generality. Let $X_n = \sum_{i=1}^k b_i \mathbb{1}_{\{X \in B_i\}}$ for b_i an arbitrary point in B_i. Then $\mathrm{E}(X_n) := \sum_{i=1}^k b_i \mathrm{P}(X \in B_i)$ satisfies $b^*(\mathrm{E}(X_n)) = \mathrm{E}(b^*(X_n))$ for all $b^* \in \mathbb{B}^*$. Also $\|\mathrm{E}(X_n) - \mathrm{E}(X_m)\| = \sup_{\|b^*\|=1} |\mathrm{E}b^*(X_n - X_m)| \leq \mathrm{E}\|X_n - X_m\| \to 0$ as $n, m \to \infty$. Thus $\mathrm{E}(X_n)$ is a Cauchy sequence in $\mathbb{B}$, and hence converges to some limit μ. Because $\mathrm{E}\|X_n - X\| < 2n^{-1}$, we have that $b^*(\mu) = \lim_{n\to\infty} b^*(\mathrm{E}X_n) = \lim_{n\to\infty} \mathrm{E}[b^*(X_n)] = \mathrm{E}[b^*(X)]$, for every $b^* \in \mathbb{B}^*$. ☐

For a mean-zero Gaussian random element W in a Banach space the RKHS is defined as the completion of the set of Pettis integrals $Sb^* = \mathrm{E}\big[b^*(W)W\big]$ with respect to the norm $\|Sb^*\|_{\mathbb{H}} = \mathrm{sd}\,\big[b^*(W)\big]$. By the Hahn-Banach theorem and the Cauchy-Schwarz inequality,

$$\begin{aligned} \|Sb^*\| &= \sup_{b_2^* \in \mathbb{B}^*: \|b_2^*\|=1} |b_2^*(Sb^*)| = \sup_{b_2^* \in \mathbb{B}^*: \|b_2^*\|=1} |\mathrm{E}[b_2^*(W) b^*(W)]| \\ &\leq \sigma(W)\ \mathrm{sd}[b^*(W)] = \sigma(W)\, \|Sb^*\|_{\mathbb{H}}. \end{aligned} \tag{I.3}$$

Thus the RKHS-norm $\|\cdot\|_{\mathbb{H}}$ on $S\mathbb{B}^*$ is stronger than the original norm $\|\cdot\|$, so that a $\|\cdot\|_{\mathbb{H}}$-Cauchy sequence in $S\mathbb{B}^* \subset \mathbb{B}$ is a $\|\cdot\|$-Cauchy sequence in $\mathbb{B}$. Consequently, the RKHS $\mathbb{H}$, which is the completion of $S\mathbb{B}^*$ under the RKHS norm, can be identified with a subset of $\mathbb{B}$. In terms of the unit balls $\mathbb{B}_1$ and $\mathbb{H}_1$ of $\mathbb{B}$ and $\mathbb{H}$ the preceding display can be written

$$\mathbb{H}_1 \subset \sigma(W)\mathbb{B}_1. \tag{I.4}$$

In other words, the norm of the embedding $\iota: \mathbb{H} \to \mathbb{B}$ is bounded by $\sigma(W)$.

Lemma I.10 *The map $S: \mathbb{B}^* \to \mathbb{H}$ is weak*-continuous.*

Proof The unit ball $\mathbb{B}_1^*$ of the dual space is weak*-metrizable (cf. Rudin 1973, 3.16). Hence the restricted map $S: \mathbb{B}_1^* \to \mathbb{H}$ is weak*- continuous if and only if weak*-convergence of a sequence b_n^* in $\mathbb{B}_1^*$ to an element b^* implies that $Sb_n^* \to Sb^*$ in $\mathbb{H}$. As the weak*-convergence $b_n^* \to b^*$ is the same as the pointwise convergence on $\mathbb{B}$, it implies that $(b_n^* - b^*)(W) \to 0$ a.s., and hence $(b_n^* - b^*)(W) \rightsquigarrow 0$. Since the latter variables are mean-zero Gaussian, the convergence is equivalent to the convergence of the variances $\|Sb_n^* - Sb^*\|_{\mathbb{H}}^2 = \mathrm{var}\,((b_n^* - b)(W)) \to 0$.

This concludes the proof that the restriction of S to the unit ball $\mathbb{B}_1^*$ is continuous. A weak*-converging net b_n^* in $\mathbb{B}^*$ is necessarily bounded in norm, by the Banach-Steinhaus theorem (Rudin 1973, 2.5), and hence is contained in a multiple of the unit ball. The continuity of the restriction then shows that $Sb_n^* \to Sb^*$, which concludes the proof. ☐

Corollary I.11 *If $\mathbb{B}_0^*$ is a weak*-dense subset of $\mathbb{B}^*$, then $\mathbb{H}$ is the completion of $S\mathbb{B}_0^*$.*

By the definitions $\langle Sb^*, S\underline{b}^*\rangle_{\mathbb{H}} = \mathrm{E}\big[b^*(W)\underline{b}^*(W)\big] = b^*(S\underline{b}^*)$, for any $b^*, \underline{b}^* \in \mathbb{B}^*$. By continuity of the inner product this extends to the *reproducing formula*: for every $h \in \mathbb{H}$ and

$b^* \in \mathbb{B}^*$,

$$\langle Sb^*, h\rangle_{\mathbb{H}} = b^*(h). \tag{I.5}$$

Read from right to left this expresses that the restriction of an element b^* of the dual space of $\mathbb{B}$ to the RKHS can be represented as an inner product.

Just as for stochastic processes there is an alternative representation of the RKHS for Banach space valued random elements through *first chaos*. In the present setting the latter is defined as the closed linear span of the variables $b^*(W)$ in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$, for $(\Omega, \mathscr{U}, P)$ the probability space on which the Gaussian process is defined. The elements Sb^* of the RKHS can be written $Sb^* = \mathrm{E}[HW]$ for $H = b^*(W)$, and the RKHS-norm of Sb^* is by definition the $\mathbb{L}_2(\Omega, \mathscr{U}, P)$-norm of this H. This immediately implies the following lemma. Note that $\mathrm{E}[HW]$ is well defined as a Pettis integral for every $H \in \mathbb{L}_2(\Omega, \mathscr{U}, P)$, by Lemma I.9.

Lemma I.12 *The RKHS of the random element W is the set of Pettis integrals* $\mathrm{E}[HW]$ *for H ranging over the closed linear span of the variables $b^*(W)$ in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$ with inner product* $\langle \mathrm{E}[H_1 W], \mathrm{E}[H_2 W]\rangle_{\mathbb{H}} = \mathrm{E}[H_1 H_2]$.

It is useful to decompose the map $S: \mathbb{B}^* \to \mathbb{B}$ as $S = A^*A$ for $A^*: \mathbb{L}_2(\Omega, \mathscr{U}, P) \to \mathbb{B}$ and $A: \mathbb{B}^* \to \mathbb{L}_2(\Omega, \mathscr{U}, P)$ given by

$$A^*H = \mathrm{E}[HW], \qquad\qquad Ab^* = b^*(W).$$

It may be checked that the operators A and A^* are indeed adjoints, after identifying $\mathbb{B}$ with a subset of its second dual space $\mathbb{B}^*$ under the canonical embedding (Rudin 1973, 3.15 and 4.5), as the notation suggests. By the preceding lemma the RKHS is the image of the first chaos space under A^*. Because $\mathrm{Range}(A)^{\perp} = \mathrm{Null}(A^*)$, the full range $A^*(\mathbb{L}_2(\Omega, \mathscr{U}, P))$ is not bigger than the image of the first chaos, but $A^*: \mathbb{L}_2(\Omega, \mathscr{U}, P) \to \mathbb{H}$ is an isometry if restricted to the first chaos space.

Recall that an operator is compact if it maps bounded sets into precompact sets, or, equivalently, maps bounded sequences into sequences that possess a converging subsequence.

Lemma I.13 *The maps $A^*: \mathbb{L}_2(\Omega, \mathscr{A}, \mathrm{P}) \to \mathbb{B}$ and $A: \mathbb{B}^* \to \mathbb{L}_2(\Omega, \mathscr{A}, \mathrm{P})$ and $S: \mathbb{B}^* \to \mathbb{B}$ are compact, and the unit ball $\mathbb{H}_1$ of the RKHS is precompact in $\mathbb{B}$.*

Proof An operator is compact if and only if its adjoint is compact, and a composition with a compact operator is compact (see Rudin 1973, 4.19). Fix some sequence b_n^* in the unit ball $\mathbb{B}_1^*$. As the unit ball is weak*- compact by the Banach-Alaoglu theorem (see Rudin 1973, 4.3(c)), there exists a subsequence along which $b_{n_j}^*$ converges pointwise on $\mathbb{B}^*$ to a limit b^*. Thus $b_{n_j}^*(W) \to b^*(W)$ a.s., and hence $\mathrm{var}\,(b_{n_j}^*(W) - b^*(W)) \to 0$. This shows that the operator A is compact.

The final assertion of the lemma follows from the fact that $\mathbb{H}_1 = A^*\mathbb{U}_1$, for $\mathbb{U}_1$ the unit ball of $\mathbb{L}_2(\Omega, \mathscr{A}, \mathrm{P})$, and hence is precompact by the compactness of A^*. □

Example I.14 (Hilbert space) The *covariance operator* of a mean-zero Gaussian random element W taking values in a Hilbert space $(\mathbb{B}, \langle\cdot,\cdot\rangle)$ is the continuous, linear, nonnegative, self-adjoint operator $S: \mathbb{B} \to \mathbb{B}$ satisfying

$$\mathrm{E}\big[\langle W, b_1\rangle\langle W, b_2\rangle\big] = \langle b_1, Sb_2\rangle, \qquad b_1, b_2 \in \mathbb{B}.$$

The RKHS of W is given by Range$(S^{1/2})$ equipped with the norm $\|S^{1/2}b\|_{\mathbb{H}} = \|b\|$. Here $S^{1/2}$ is the square root of S: the positive, self-adjoint operator $S^{1/2}: \mathbb{B} \to \mathbb{B}$ such that $S^{1/2}S^{1/2} = S$.

To prove this, observe first that the covariance operator S coincides with the operator S defined by the Pettis integral $Sb^* = \mathrm{E}\big[b^*(W)W\big]$ under the natural identification of $\mathbb{B}^*$ with $\mathbb{B}$ (where $b \in \mathbb{B}$ corresponds to the element $b_1 \mapsto \langle b, b_1\rangle$ of $\mathbb{B}^*$). Hence the RKHS is the completion of $\{Sb: b \in \mathbb{B}\}$ under the norm $\|Sb\|_{\mathbb{H}} = \mathrm{sd}[\langle W, b\rangle] = \|S^{1/2}b\|$. This is the same as the completion of $\{S^{1/2}c: c \in S^{1/2}\mathbb{B}\}$ under the norm $\|S^{1/2}c\|_{\mathbb{H}} = \|c\|$. The latter set is already complete, so that the completion operation is superfluous.

Example I.15 ($\mathbb{L}_2$-space) A measurable stochastic process W with $\int_0^1 W_s^2\,ds < \infty$, for every sample path, is a random element in $\mathbb{L}_2[0,1]$. The dual space of $\mathbb{L}_2[0,1]$ consists of the maps $g \mapsto \int g(s)f(s)\,ds$ for f ranging over $\mathbb{L}_2[0,1]$. By Fubini's theorem, with K the covariance function of W,

$$Sf(t) = \mathrm{E}\Big[W_t \int_0^1 W_s f(s)\,ds\Big] = \int_0^1 K(s,t) f(s)\,ds.$$

Thus S coincides with the kernel operator with kernel K. The RKHS is the completion of the range of K under the inner product $\langle Sf, Sg\rangle_{\mathbb{H}} = \int_0^1\int_0^1 K(s,t)f(s)g(t)\,ds dt$.

RKHS under Transformation

The image of a Gaussian random element under a continuous, linear map is again a Gaussian random element. If the map is also one-to-one, then the RKHS is transformed in parallel.

Lemma I.16 *Let W be a mean-zero Gaussian random element in $\mathbb{B}$ with RKHS $\mathbb{H}$. Let $T: \mathbb{B} \to \underline{\mathbb{B}}$ be a one-to-one, continuous, linear map from $\mathbb{B}$ into another Banach space $\underline{\mathbb{B}}$. Then the RKHS of the Gaussian random element TW in $\underline{\mathbb{B}}$ is given by $T\mathbb{H}$ and $T: \mathbb{H} \to \underline{\mathbb{H}}$ is a Hilbert space isometry.*

Proof Let $T^*: \underline{\mathbb{B}}^* \to \mathbb{B}^*$ be the adjoint of T, so that $(T^*\underline{b}^*)(b) = \underline{b}^*(Tb)$, for every $\underline{b}^* \in \underline{\mathbb{B}}^*$ and $b \in \mathbb{B}$. The RKHS $\underline{\mathbb{H}}$ of TW is the completion of $\{\underline{S}\underline{b}^*: \underline{b}^* \in \underline{\mathbb{B}}^*\}$, where

$$\underline{S}\underline{b}^* = \mathrm{E}\big[(TW)\underline{b}^*(TW)\big] = T\Big(\mathrm{E}\big[W\underline{b}^*(TW)\big]\Big) = T\Big(\mathrm{E}\big[W(T^*\underline{b}^*)(W)\big]\Big) = TST^*\underline{b}^*.$$

Furthermore, the inner product in $\underline{\mathbb{H}}$ is given by

$$\langle \underline{S}\underline{b}_1^*, \underline{S}\underline{b}_2^*\rangle_{\underline{\mathbb{H}}} = \mathrm{E}\big[\underline{b}_1^*(TW)\underline{b}_2^*(TW)\big] = \mathrm{E}\big[(T^*\underline{b}_1^*W)(T^*\underline{b}_2^*W)\big] = \langle ST^*\underline{b}_1^*, ST^*\underline{b}_2^*\rangle_{\mathbb{H}}.$$

Hence it follows that $\underline{S}\underline{b}^* = T(ST^*\underline{b}^*)$, and $\|\underline{S}\underline{b}^*\|_{\underline{\mathbb{H}}} = \|ST^*\underline{b}^*\|_{\mathbb{H}}$. Thus the linear map $T: ST^*\underline{\mathbb{B}}^* \subset \mathbb{H} \to \underline{S}\underline{\mathbb{B}}^* \subset \underline{\mathbb{H}}$ is an isometry for the RKHS-norms. It extends by continuity to a linear map from the completion $\mathbb{H}_0$ of $ST^*\underline{\mathbb{B}}^*$ in $\mathbb{H}$ to $\underline{\mathbb{H}}$. Because T is continuous for the norms of $\mathbb{B}$ and $\underline{\mathbb{B}}$ and the RKHS-norms are stronger, this extension agrees with T. Since $\mathbb{H}_0$ and $\underline{\mathbb{H}}$ are, by definition, the completions of $ST^*\underline{\mathbb{B}}^*$ and $\underline{S}\underline{\mathbb{B}}^*$, we have that $T: \mathbb{H}_0 \to \underline{\mathbb{H}}$ is an isometry onto $\underline{\mathbb{H}}$. It remains to show that $\mathbb{H}_0 = \mathbb{H}$.

Because T is one-to-one, its range $T^*(\underline{\mathbb{B}}^*)$ is weak*- dense in $\mathbb{B}^*$; see Rudin (1973), Corollary 4.12. In view of Lemma I.10, the map $S: \mathbb{B}^* \to \mathbb{H}$ is continuous relative to the weak*- and RKHS topologies. Thus $S(T^*\underline{\mathbb{B}}^*)$ is dense in $S\mathbb{B}^*$ for the RKHS-norm of $\mathbb{H}$ and hence is dense in $\mathbb{H}$. Thus $\mathbb{H}_0 = \mathbb{H}$. □

RKHS Relative to Different Norms

A stochastic process W can often be viewed as a map into several Banach spaces. For instance, a process indexed by the unit interval with continuous sample paths is a Borel measurable map in $\mathfrak{C}[0, 1]$, as well as in $\mathbb{L}_2[0, 1]$. A process with continuously differentiable sample paths is a map in $\mathfrak{C}[0, 1]$, in addition to being a map in $\mathfrak{C}^\alpha[0, 1]$, for $\alpha < 1$. The RKHS obtained from using a weaker Banach space norm (corresponding to a continuous embedding in a bigger Banach space) is typically the same. One could say that RKHS is *intrinsic* to the process.

Lemma I.17 *Let* $(\mathbb{B}, \|\cdot\|)$ *be a separable Banach space and let* $\|\cdot\|'$ *be a norm on* $\mathbb{B}$ *with* $\|b\|' \leq \|b\|$. *Then the RKHS of a Gaussian random element* W *in* $(\mathbb{B}, \|\cdot\|)$ *is the same as the RKHS of* W *viewed as a random element in the completion of* $\mathbb{B}$ *under* $\|\cdot\|'$.

Proof Let $\mathbb{B}'$ be the completion of $\mathbb{B}$ relative to $\|\cdot\|'$. The assumptions imply that the identity map $\iota: (\mathbb{B}, \|\cdot\|) \to (\mathbb{B}', \|\cdot\|')$ is continuous, linear and one-to-one. Hence the proposition follows from Lemma I.16. □

RKHS under Independent Sum

If a given Gaussian prior misses certain desirable "directions" in its RKHS, then these can be filled in by adding independent Gaussian components in these directions.

Recall that a closed linear subspace $\mathbb{B}_0 \subset \mathbb{B}$ of a Banach space $\mathbb{B}$ is *complemented* if there exists a closed linear subspace $\mathbb{B}_1$ with $\mathbb{B} = \mathbb{B}_0 + \mathbb{B}_1$ and $\mathbb{B}_0 \cap \mathbb{B}_1 = \{0\}$. All closed subspaces of a Hilbert space are complemented, but in a general Banach space this is not the case. However, finite-dimensional subspaces and subspaces that are the full space up to a finite-dimensional space are complemented in every Banach space, as a consequence of the Hahn-Banach theorem.

Lemma I.18 *Let* V *and* W *be independent mean-zero Gaussian random elements taking values in subspaces* $\mathbb{B}^V$ *and* $\mathbb{B}^W$ *of a separable Banach space* $\mathbb{B}$ *with RKHSs* $\mathbb{H}^V$ *and* $\mathbb{H}^W$ *respectively. Assume that* $\mathbb{B}^V \cap \mathbb{B}^W = \{0\}$ *and that* $\mathbb{B}^V$ *is complemented in* $\mathbb{B}$ *by a subspace that contains* $\mathbb{B}^W$. *Then the RKHS of* $V + W$ *is given by the direct sum* $\mathbb{H}^V \oplus \mathbb{H}^W$ *and the RKHS norms satisfy* $\|h^V + h^W\|^2_{\mathbb{H}^{V+W}} = \|h^V\|^2_{\mathbb{H}^V} + \|h^W\|^2_{\mathbb{H}^W}$.

Proof By the independence of V and W, for any $b^* \in \mathbb{B}^*$,

$$S^{V+W} b^* = \mathrm{E}\big[b^*(V + W)(V + W)\big] = S^V b^* + S^W b^*.$$

The assumptions that $\mathbb{B}^V \cap \mathbb{B}^W = \{0\}$ and $\mathbb{B}^V$ is complemented by a subspace that contains $\mathbb{B}^W$ imply that there exists a continuous linear map $\Pi: \mathbb{B} \to \mathbb{B}^V$ such that $\Pi b = b$ if $b \in \mathbb{B}^V$ and $\Pi b = 0$ if $b \in \mathbb{B}^W$. (This is a consequence of the Hahn-Banach Theorem [cf.

Theorem 3.2 of Rudin 1973].) Then $\Pi V = V$ and $(I-\Pi)W = W$ and $(I-\Pi)V = \Pi W = 0$ a.s., which can be seen to imply, for every $b^* \in \mathbb{B}^*$,

$$S^V b^* = S^V(b^*\Pi), \qquad S^W b^* = S^W(b^*(I-\Pi)), \qquad S^V(b^*(I-\Pi)) = S^W(b^*\Pi) = 0.$$

Given $b_1^*, b_2^* \in \mathbb{B}^*$ the map $b^* = b_1^*\Pi + b_2^*(I-\Pi)$ is also an element of $\mathbb{B}^*$, and $S^{V+W}b^* = S^V b_1^* + S^W b_2^*$. It is also seen that $\|S^{V+W}b^*\|^2_{\mathbb{H}^{V+W}} = \mathrm{var}[b^*(V+W)] = \mathrm{var}[(b_1^*(V) + b_2^*(W)]$ is the sum of $\|S^V b^*\|^2_{\mathbb{H}^V}$ and $\|S^W b^*\|^2_{\mathbb{H}^W}$. ☐

The assumption that $\mathbb{B}^V \cap \mathbb{B}^W = \{0\}$ can be interpreted as requiring "linear independence" rather than some form of orthogonality of the supports of V and W. The lemma shows that stochastic independence of V and W translates the linear independence into orthogonality in the RKHS of $V+W$.

The lemma assumes trivial intersection of the *supports* of the variables V and W, not merely trivial intersection of linear subspaces containing the ranges of the variables. As the RKHS is independent of the norm in view of Lemma I.17, the closure may be taken with respect to the strongest possible norm, to make the supports as small as possible and enhance the possibility of a trivial intersection.

The assumption of trivial intersection cannot be removed, as the following example shows.

Example I.19 Let $Z_{i1}, Z_{i2} \overset{\text{iid}}{\sim} \mathrm{Nor}(0,1)$, for $i = 1, 2, \ldots$, and define two independent Gaussian processes by the series $V_j = \sum_{i=1}^\infty \mu_{ij} Z_{ij}\psi_i$, for $j = 1, 2$, where $\{\psi_i\}$ is a basis in some Banach space. Then

$$V_1 + V_2 = \sum_{i=1}^\infty (\mu_{i1}Z_{i1} + \mu_{i2}Z_{i2})\psi_i = \sum_{i=1}^\infty \mu_i Z_i \psi_i,$$

for $\mu_i^2 = \mu_{i1}^2 + \mu_{i2}^2$ and $Z_i \overset{\text{iid}}{\sim} \mathrm{Nor}(0,1)$, for $i = 1, 2, \ldots$. As shown in Section I.6, the RKHS of $V_1 + V_2$ consists of the series $\sum_{i=1}^\infty w_i\psi_i$ with $\sum_{i=1}^\infty (w_i^2/\mu_i^2) < \infty$. The RKHSs of V_1 and V_2 are characterized similarly, and in general the RKHS of $V_1 + V_2$ is not an orthogonal sum of the latter RKHSs. In fact, the RKHS of $V_1 + V_2$ depends crucially on the order at which μ_i tends to as $i \to \infty$, and this is determined by the larger of μ_{i1} and μ_{i2}. If $\mu_{i1}/\mu_{i2} \to 0$, then the RKHS of $V_1 + V_2$ essentially coincides with the RKHS of V_2, but the distributions of V_1 and V_2 are orthogonal in that case.

I.5 Absolute Continuity

For a mean-zero Gaussian random element W in a separable Banach space $\mathbb{B}$ defined on a probability space $(\Omega, \mathscr{U}, P)$ and $\mathbb{H}$ its RKHS, define a map U by

$$U(Sb^*) = b^*(W), \qquad b^* \in \mathbb{B}^*. \tag{I.6}$$

By the definition of the RKHS the map $S\mathbb{B}^*: \mathbb{H} \to \mathbb{L}_2(\Omega, \mathscr{U}, P)$ is an isometry. Let $U: \mathbb{H} \to \mathbb{L}_2(\Omega, \mathscr{U}, P)$ be its extension to the full RKHS.

Proposition I.20 *If W is a mean-zero Gaussian random element in a separable Banach space and h is an element of its RKHS, then the distributions P^{W+h} and P^W of $W+h$ and W on $\mathbb{B}$ are equivalent with Radon-Nikodym density*

$$\frac{dP^{W+h}}{dP^W}(W) = e^{Uh-\frac{1}{2}\|h\|_{\mathbb{H}}^2}, \qquad \text{a.s.}$$

Proof Because Uh is in the closed linear span of the mean-zero Gaussian variables $b^*(W)$, it is itself mean-zero Gaussian variable. By the isometry property of U its variance is equal to $\mathrm{var}[Uh] = \|h\|_{\mathbb{H}}^2$, and hence

$$dQ = e^{Uh-\frac{1}{2}\|h\|_{\mathbb{H}}^2}\, dP$$

defines a probability measure on $(\Omega, \mathscr{U})$. For any pair $b_1^*, b_2^* \in \mathbb{B}^*$ the joint distribution of the random vector $(USb_1^*, USb_2^*) = (b_1^*W, b_2^*W)$ under P is bivariate normal with mean zero and covariance matrix $(\langle Sb_i^*, Sb_j^*\rangle_{\mathbb{H}})_{i,j=1,2}$. By taking limits we see that for every $h \in \mathbb{H}$ the joint distribution of the vector (b_1^*W, Uh) is bivariate normal with mean zero and covariance matrix Σ with $\Sigma_{1,1} = \|Sb_1^*\|_{\mathbb{H}}^2$, $\Sigma_{1,2} = \langle Sb_1^*, h\rangle_{\mathbb{H}}$ and $\Sigma_{2,2} = \|h\|_{\mathbb{H}}^2$. Thus

$$\begin{aligned} \mathrm{E}_Q e^{ib_1^*(W)} &= \mathrm{E} e^{ib_1^*(W)} e^{Uh-\frac{1}{2}\|h\|_{\mathbb{H}}^2} = e^{\frac{1}{2}(i,1)\Sigma(i,1)^\top} e^{-\frac{1}{2}\|h\|_{\mathbb{H}}^2} = e^{-\frac{1}{2}\Sigma_{1,1}+i\Sigma_{1,2}} \\ &= \mathrm{E} e^{ib_1^*W + i\langle Sb_1^*, h\rangle_{\mathbb{H}}}. \end{aligned}$$

By the reproducing formula (I.5) we have $\langle Sb_1^*, h\rangle_{\mathbb{H}} = b_1^*(h)$, whence the right side is equal to $\mathrm{E} e^{ib_1^*(W+h)}$. From uniqueness of characteristic functions we conclude that the distribution of $b^*(W)$ under Q is the same as the distribution of $b^*(W+h)$ under P, for every $b^* \in \mathbb{B}^*$. This implies that the distribution of $W+h$ under P is the same as the distribution of W under Q, i.e. $P(W+h \in B) = \mathrm{E}_Q \mathbb{1}_B(W) = \mathrm{E}\mathbb{1}_B(W)(dQ/dP)$, which shows that dQ/dP is the desired Radon-Nikodym derivative. □

The preceding lemma requires that the shift h is contained in the RKHS. If this is not the case, then the two Gaussian measures are orthogonal and hence there is no density (see e.g. van der Vaart and van Zanten 2008b, Lemma 3.3 for a proof).

I.6 Series Representation

Consider a covariance kernel K of a Gaussian process $W = (W_t: t \in T)$ of the form, for given $\lambda_1, \lambda_2, \ldots > 0$ and arbitrary functions $\phi_j: T \to \mathbb{R}$,

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(t). \tag{I.7}$$

The series is assumed to converge pointwise on $T \times T$; the index set T may be arbitrary. The following theorem characterizes the stochastic process RKHS, under the condition that the functions ϕ_j are *infinitely linearly independent* in the following sense: if $\sum_{j=1}^\infty w_j\phi_j(t) = 0$ for every $t \in T$ and some sequence w_j with $\sum_{j=1}^\infty w_j^2\lambda_j^{-1} < \infty$, then $w_j = 0$ for every $j \in \mathbb{N}$.

Theorem I.21 (RKHS series) *If the covariance function K of a mean-zero Gaussian stochastic process $W = (W_t: t \in T)$ can be represented as in* (I.7) *for positive numbers λ_j and infinitely linearly independent functions $\phi_j: T \to \mathbb{R}$ such that $\sum_{j=1}^\infty \lambda_j \phi_j^2(t) < \infty$ for every $t \in T$, then the RKHS of W consists of all functions of the form $\sum_{j=1}^\infty w_j \phi_j$ with $\sum_{j=1}^\infty w_j^2 \lambda_j^{-1} < \infty$ with inner product*

$$\Big\langle \sum_{j=1}^\infty v_j \phi_j, \sum_{j=1}^\infty w_j \phi_j \Big\rangle_{\mathbb{H}} = \sum_{j=1}^\infty \frac{v_j w_j}{\lambda_j}.$$

Proof By the Cauchy-Schwarz inequality and the assumption that $\sum_{j=1}^\infty \lambda_j \phi_j^2(t) < \infty$ for every $t \in T$, the series (I.7) converges absolutely for every $(s,t) \in T \times T$. The same is true for every series $f_w(t) := \sum_{j=1}^\infty w_j \phi_j(t)$ with coefficients such that $(w_j/\lambda_j^{-1/2}) \in \ell_2$. Thus $f_w: T \to \mathbb{R}$ defines a function, which by the assumption of linear independence of the basis functions corresponds to a unique sequence of coefficients (w_j). The inner product $\langle f_v, f_w \rangle_{\mathbb{H}}$ as given in the theorem gives an isometry between the functions and the sequences $(w_j/\lambda_j^{-1/2}) \in \ell_2$ and hence the set of all functions f_w is a Hilbert space H under this inner product.

It suffices to show that H coincides with the RKHS. By (I.7) for every $s \in T$ the function $t \mapsto K(s,t)$ is contained in H with coefficients $w_j = \lambda_j \phi_j(s)$. Furthermore, for $s, t \in T$,

$$\langle K(s,\cdot), K(t,\cdot) \rangle_{\mathbb{H}} = \sum_{k=1}^\infty \frac{\lambda_k \phi_k(s) \lambda_k \phi_k(t)}{\lambda_k} = K(s,t).$$

By Definition 11.12 and the following discussion, this shows that on the linear span of the functions $K(s,\cdot)$ the given inner product indeed coincides with the RKHS inner product. The RKHS is the completion of this linear span by definition, whence it suffices to show that H is not bigger than the RKHS. For $t \in T$ and any $f_w \in H$,

$$\langle f_w, K(t,\cdot) \rangle_{\mathbb{H}} = \Big\langle \sum_{j=1}^\infty w_j \phi_j, \sum_{j=1}^\infty \lambda_j \phi_j(t) \phi_j \Big\rangle_{\mathbb{H}} = \sum_{j=1}^\infty \frac{w_j \lambda_j \phi_j(t)}{\lambda_j} = f_w(t).$$

If $f_w \in H$ with $f_w \perp \mathbb{H}$, then in particular $f_w \perp K(t,\cdot)$ for every $t \in T$. Then the preceding reproducing formula show that $f_w(t) = 0$, for all t. Hence H is equal to the RKHS. $\square$

Series expansions of the type (I.7) are not unique, and some may be more useful than others. They may arise as an eigenvalue expansion of the operator corresponding to the covariance function. However, this is not a requirement of the proposition, which applies to arbitrary functions ϕ_j.

Example I.22 (Eigen expansion) Suppose that $(T, \mathscr{T}, \nu)$ is a measure space and $K: T \times T \to \mathbb{R}$ is a covariance kernel such that $\iint K^2(s,t)\, d\nu(s)\, d\nu(t) < \infty$. Then the integral operator $K: \mathbb{L}_2(T, \mathscr{T}, \nu) \to \mathbb{L}_2(T, \mathscr{T}, \nu)$ defined by

$$Kf(t) = \int f(s)\, K(s,t)\, d\nu(t)$$

is compact and positive self-adjoint. Then there exists a sequence of eigenvalues $\lambda_j \downarrow 0$ and an orthonormal system of eigenfunctions $\phi_j \in \mathbb{L}_2(T, \mathscr{T}, \nu)$ (thus $K\phi_j = \lambda_j \phi_j$ for every $j \in \mathbb{N}$) such that (I.7) holds, except that the series converges in $\mathbb{L}_2(T \times T, \mathscr{T} \times \mathscr{T}, \nu \times \nu)$. The series $\sum_j w_j \phi_j$ now converges in $\mathbb{L}_2(T, \mathscr{T}, \nu)$ for any sequence (w_j) in ℓ_2. By the orthonormality of the functions ϕ_j, they are certainly linearly independent.

If the series (I.7) also converges pointwise on $T \times T$, then in particular $K(t,t) = \sum_j \lambda_j \phi_j^2(t) < \infty$ for all $t \in T$ and Theorem I.21 shows that the RKHS is the set of all functions $\sum_k w_k \phi_k$ for sequences (w_j) such that $(w_j/\lambda_j^{-1/2}) \in \ell_2$.

If the kernel is suitably regular, then we can apply the preceding with many choices of measure ν, leading to different eigenfunction expansions.

If the series (I.7) does not converge pointwise, then the preceding theorem does not apply. However, by Example I.15 the RKHS can be characterized as the range of the operator K with square norm $\|Kf\|_{\mathbb{H}}^2 = \int (Kf) f \, d\nu$. Since $f = \sum_j f_j \phi_j$ for $(f_j) \in \ell_2$ and $Kf = \sum_j \lambda_j f_j \phi_j$ and $\int (Kf) f \, d\nu = \sum_j \lambda_j f_j^2$, this shows that the analogous result is still true (make the substitution $\lambda_j f_j = w_j$).

Consider the stochastic process of the form, for a sequence of numbers μ_j, i.i.d. standard normal variables (Z_j) and suitable functions ϕ_j,

$$W_t = \sum_{j=1}^{\infty} \mu_j Z_j \phi_j(t).$$

If this series converges in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$ for every t, then (I.7) holds with $\lambda_j = \mu_j^2$. The stochastic process RKHS then takes the form given by the preceding proposition.

The following theorem gives a Banach space version of this result. Say that a sequence (h_j) of elements of a separable Banach space is *linearly independent* over ℓ_2 if $\sum_{j=1}^{\infty} w_j h_j = 0$ for for some $w \in \ell_2$, where the convergence of the series is in $\mathbb{B}$, implies that $w = 0$.

Theorem I.23 (RKHS series) *If for a given sequence (h_j) of ℓ_2-linearly independent elements of a separable Banach space $\mathbb{B}$ and a sequence (Z_j) of i.i.d. standard normal variables the series $W = \sum_{j=1}^{\infty} Z_j h_j$ converges almost surely in $\mathbb{B}$, then the RKHS of W as a map in $\mathbb{B}$ is the set of all elements $\sum_{i=1}^{\infty} w_j h_j$ for $w \in \ell_2$ with squared norm $\|\sum_j w_j h_j\|_{\mathbb{H}}^2 = \sum_j w_j^2$.*

Proof The almost sure convergence of the series $W = \sum_j Z_j h_j$ in $\mathbb{B}$ implies the almost sure convergence of the series $b^*(W) = \sum_j Z_j b^*(h_j)$ in $\mathbb{R}$, for any $b^* \in \mathbb{B}^*$. Because the partial sums of the last series are mean-zero Gaussian, the series converges also in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$. Hence for any $b^*, \underline{b}^* \in \mathbb{B}^*$,

$$\mathrm{E}\big[b^*(W)\underline{b}^*(W)\big] = \mathrm{E}\Big[\sum_{j=1}^{\infty} Z_j b^*(h_j) \sum_{j=1}^{\infty} Z_j \underline{b}^*(h_j)\Big] = \sum_{j=1}^{\infty} b^*(h_j)\underline{b}^*(h_j).$$

In particular, the sequence $(b^*(h_j))_{j=1}^\infty$ is contained in ℓ_2 for every $b^* \in \mathbb{B}^*$, with square norm $\mathrm{E}\big[b^*(W)^2\big]$.

For $w \in \ell_2$ and natural numbers $m < n$, by the Hahn-Banach theorem and the Cauchy-Schwarz inequality,

$$\Big\|\sum_{m<j\le n} w_j h_j\Big\|^2 = \sup_{\|b^*\|\le 1}\Big|\sum_{m<j\le n} w_j b^*(h_j)\Big|^2 \le \sum_{m<j\le n} w_j^2 \sup_{\|b^*\|\le 1}\sum_{m<j\le n} (b^*(h_j))^2.$$

As $m,n \to \infty$ the first factor on the far right tends to zero, since $w \in \ell_2$. By the first paragraph the second factor is bounded by $\sup_{\|b^*\|\le 1}\mathrm{E}\big[b^*(W)^2\big] \le \mathrm{E}\|W\|^2$. Hence the partial sums of the series $\sum_j w_j h_j$ form a Cauchy sequence in $\mathbb{B}$, whence the infinite series converges.

Because the sequence $w_j = b^*(h_j)$ is contained in ℓ_2, the series $\sum_j b^*(h_j)h_j$ converges in $\mathbb{B}$, and hence $\underline{b}^*\big(\sum_j b^*(h_j)h_j\big) = \sum_j b^*(h_j)\underline{b}^*(h_j) = \mathrm{E}\big[b^*(W)\underline{b}^*W\big]$, for any $\underline{b}^* \in \mathbb{B}^*$. This shows that $Sb^* = \sum_j (b^*h_j)h_j$ and hence the RKHS is not bigger than the space as claimed.

The space would be smaller than claimed if there existed $w \in \ell_2$ that is not in the closure of the linear span of the elements $b^*(h_j)$ of ℓ_2 when b^* ranges over $\mathbb{B}^*$. We can take this w without loss of generality orthogonal to the latter collection, i.e. $\sum_j w_j b^*(h_j) = 0$ for every $b^* \in \mathbb{B}^*$. This is equivalent to $\sum_j w_j h_j = 0$, but this is excluded for any $w \neq 0$ by the assumption of linear independence of the h_j. □

The sequence (h_j) in the preceding theorem may consist of arbitrary elements of the Banach space, only restricted by linear independence over ℓ_2 and the convergence of the random sequence $\sum_j Z_j h_j$. The theorem shows that when combined in a series with i.i.d. standard normal coefficients, then this sequence turns into an *orthonormal* basis of the RKHS.

From the proof it can be seen that the linear independence is necessary; see Problem I.2.

Example I.24 (Polynomials) For $Z_0, \ldots, Z_k$ i.i.d. standard normal variables consider the polynomial process $t \mapsto \sum_{j=0}^k Z_j t^j/j!$ viewed as a map in (for instance) $\mathfrak{C}[0,1]$. The RKHS of this process is equal to the set of kth degree polynomials $P_a(t) = \sum_{j=0}^k a_j t^j/j!$ with square norm $\|P_a\|_{\mathbb{H}}^2 = \sum_{j=0}^k a_j^2$. In other words, the kth degree polynomials P with square norm $\|P\|_{\mathbb{H}}^2 = \sum_{j=0}^k P^{(j)}(0)^2$.

The following theorem shows that, conversely, any Gaussian random element W in a separable Banach space can be expanded in a series $W = \sum_{j=1}^\infty Z_j h_j$, for i.i.d. standard normal variables Z_j and any orthonormal basis (h_j) of its RKHS, where the series converges in the norm of the Banach space. Because we can rewrite this expansion as $W = \sum_j \|h_j\| Z_j \tilde{h}_j$, where $\tilde{h}_j = h_j/\|h_j\|$ is a sequence of norm one, the corresponding "eigenvalues" λ_j are in this case the square norms $\|h_j\|^2$.

To formulate the theorem, recall the isometry $U\colon \mathbb{H} \to \mathbb{L}_2(\Omega, \mathscr{U}, P)$ defined by $U(Sb^*) = b^*(W)$ in (I.6).

Theorem I.25 (Series representation) *Let* (h_j) *be a complete orthonormal system in the RKHS of a Borel measurable, mean-zero Gaussian random element* W *in a separable Banach space* $\mathbb{B}$*. Then* $Uh_1, Uh_2, \ldots$ *is an i.i.d. sequence of standard normal variables and* $W = \sum_{i=1}^{\infty}(Uh_j)h_j$*, where the series converges in the norm of* $\mathbb{B}$*, almost surely.*

Proof It is immediate from the definitions of U and the RKHS that $U\colon \mathbb{H} \to \mathbb{L}_2(\Omega, U, P)$ is an isometry. Because U maps the subspace $S\mathbb{B}^* \subset \mathbb{H}$ into the Gaussian process $b^*(W)$ indexed by $b^* \in \mathbb{B}^*$, it maps the completion $\mathbb{H}$ of $S\mathbb{B}^*$ into the completion of the linear span of this process in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$, which consists of normally distributed variables. Because U retains inner products, it follows that $Uh_1, Uh_2, \ldots$ is a sequence of i.i.d. standard normal variables.

If $W_n = \sum_{j=1}^{n}(Uh_j)h_j$, then $\mathrm{E}(W_n|\, Uh_1, \ldots, Uh_m) = W_m$, for every $m \leq n$, in a Banach space sense. Convergence of the infinite series follows by a martingale convergence theorem for Banach space valued variables; see Ledoux and Talagrand (1991), Proposition 3.6. $\square$

I.7 Support and Concentration

In this section we provide proofs of the key concentration lemmas Lemma 11.18 and 11.19.

In Chapter 11 the *concentration function* of the Gaussian random element W is defined by

$$\varphi_w(\epsilon) = \inf_{h \in \mathbb{H}: \|h-w\| \leq \epsilon} \tfrac{1}{2}\|h\|_{\mathbb{H}}^2 - \log \mathrm{P}(\|W\| < \epsilon).$$

Lemma I.26 *For any* w *in* $\mathbb{B}$ *the concentration function* $\epsilon \mapsto \varphi_w(\epsilon)$ *of a nondegenerate mean-zero Gaussian random element in a separable Banach space is strictly decreasing and convex on* $(0, \infty)$*, and hence continuous.*

Proof The centered small ball exponent φ_0 is convex by Lemma 1.1 of Gaenssler et al. (2007). The decentering function (the infimum) is also convex, as a consequence of the convexity of the norms of $\mathbb{B}$ and $\mathbb{H}$.

The decentering function is clearly nondecreasing. We show that the centered small ball exponent is strictly decreasing by showing that $\mathrm{P}(\epsilon < \|W\| < \epsilon') > 0$, whenever $0 < \epsilon < \epsilon'$. Since W is nondegenerate, the RKHS $\mathbb{H}$ contains some nonzero element h, which we can scale so that $\epsilon < \|h\| < \epsilon'$. Then a ball of sufficiently small radius centered at h is contained in $\{b \in \mathbb{B}\colon \epsilon < \|b\| < \epsilon'\}$ and has positive probability, as the RKHS belongs to the support of W. $\square$

Lemma I.27 *For every* h *in the RKHS of mean-zero Gaussian random element* W *in a separable Banach space* $\mathbb{B}$ *and every Borel measurable set* $C \subset \mathbb{B}$ *with* $C = -C$*,*

$$\mathrm{P}(W - h \in C) \geq e^{-\frac{1}{2}\|h\|_{\mathbb{H}}^2}\mathrm{P}(W \in C).$$

Proof Since $W =_d -W$ and $C = -C$ we have $\mathrm{P}(W + h \in C) = \mathrm{P}(-W + h \in -C) = \mathrm{P}(W - h \in C)$. By Lemma I.20,

$$\mathrm{P}(W + h \in C) = \mathrm{E}[\mathbb{1}_C(W + h)] = \mathrm{E}[e^{Uh - \frac{1}{2}\|h\|_{\mathbb{H}}^2}\mathbb{1}_C(W)].$$

Since the left side remains the same if h is replaced $-h$,

$$\mathrm{P}(W - h \in C) = \tfrac{1}{2}\mathrm{E}[e^{Uh - \frac{1}{2}\|h\|_{\mathbb{H}}^2}\mathbb{1}_C(W)] + \tfrac{1}{2}\mathrm{E}[e^{U(-h) - \frac{1}{2}\|-h\|_{\mathbb{H}}^2}\mathbb{1}_C(W)].$$

The result follows since $(e^x + e^{-x})/2 \geq 1$, for every x. □

The lemma with $C = \epsilon\mathbb{B}_1$ relates the decentered small ball probability $\mathrm{P}(\|W - h\| < \epsilon)$ to the corresponding centered small ball probability, for $h \in \mathbb{H}_1$. The following lemma extends to a ball around a general element of $\mathbb{B}$.

Lemma I.28 *For every w in the closure of the RKHS of a mean-zero Gaussian random element W in a separable Banach space $\mathbb{B}$, and every $\epsilon > 0$,*

$$\varphi_w(\epsilon) \leq -\log \mathrm{P}(\|W - w\| < \epsilon) \leq \varphi_w(\epsilon/2).$$

Proof If $h \in \mathbb{H}$ is such that $\|h - w\| \leq \epsilon$, then $\|W - w\| \leq \epsilon + \|W - h\|$ by the triangle inequality and hence

$$\mathrm{P}(\|W - w\| < 2\epsilon) \geq \mathrm{P}(\|W - h\| < \epsilon) \geq e^{-\|h\|_{\mathbb{H}}^2/2}\mathrm{P}(\|W\| < \epsilon),$$

by Lemma I.27. Taking the negative logarithm and optimizing over $h \in \mathbb{H}$, we obtain the upper bound of the lemma.

The set $B_\epsilon = \{h \in \mathbb{H}\colon \|h - w\| \leq \epsilon\}$ is convex, and closed in $\mathbb{H}$, because the RKHS norm is stronger than the Banach space norm. Thus the convex map $h \mapsto \|h\|_{\mathbb{H}}^2$ attains a minimum on B_ϵ at some point h_ϵ. Because $(1 - \lambda)h_\epsilon + \lambda h \in B_\epsilon$ for every $h \in B_\epsilon$ and $0 \leq \lambda \leq 1$, we have $\|(1 - \lambda)h_\epsilon + \lambda h\|_{\mathbb{H}}^2 \geq \|h_\epsilon\|_{\mathbb{H}}^2$, whence $2\lambda\langle h - h_\epsilon, h_\epsilon\rangle_{\mathbb{H}} + \lambda^2\|h - h_\epsilon\|_{\mathbb{H}}^2 \geq 0$. Since $0 \leq \lambda \leq 1$ can be arbitrary, this gives $\langle h - h_\epsilon, h_\epsilon\rangle_{\mathbb{H}} \geq 0$, or

$$\langle h, h_\epsilon\rangle_{\mathbb{H}} \geq \|h_\epsilon\|_{\mathbb{H}}^2 \text{ for every } h \in B_\epsilon.$$

By Theorem I.25 the Gaussian element can be represented as $W = \sum_{i=1}^\infty (Uh_j)h_j$, where the convergence is in the norm of $\mathbb{B}$, almost surely, and where $\{h_1, h_2, \ldots\}$ is a complete orthonormal system for $\mathbb{H}$. The variable $W_m := \sum_{i=1}^m (Uh_j)h_j$ takes its values in $\mathbb{H}$, and for any $g \in \mathbb{H}$ satisfies $\|W_m - g - w\| < \epsilon$ for sufficiently large m, a.s. on the event $\|W - g - w\| < \epsilon$. In other words, $W_m - g \in B_\epsilon$ eventually a.s. on the event $\|W - g - w\| < \epsilon$, and hence by the preceding display $\langle W_m - g, h_\epsilon\rangle_{\mathbb{H}} \geq \|h_\epsilon\|_{\mathbb{H}}^2$ for all sufficiently large m a.s. Since $\langle W_m, h_\epsilon\rangle_{\mathbb{H}} = \sum_{i=1}^m (Uh_j)\langle h_j, h_\epsilon\rangle_{\mathbb{H}} = U\sum_{i=1}^m h_j\langle h_j, h_\epsilon\rangle_{\mathbb{H}}$, the sequence $\langle W_m, h_\epsilon\rangle_{\mathbb{H}}$ tends to Uh_ϵ as $m \to \infty$, in $\mathbb{L}_2(\Omega, \mathscr{U}, P)$ and hence also almost surely along a subsequence. We conclude that $Uh_\epsilon - \langle g, h_\epsilon\rangle_{\mathbb{H}} \geq \|h_\epsilon\|_{\mathbb{H}}^2$ a.s. on the event $\|W - g - w\| < \epsilon$. For $g = -h_\epsilon$, this gives that $Uh_\epsilon \geq 0$ a.s. on the event $\|W + h_\epsilon - w\| < \epsilon$. By Lemma I.20,

$$\begin{aligned}\mathrm{P}(W \in w + \epsilon\mathbb{B}_1) &= \mathrm{P}(W - h_\epsilon \in w - h_\epsilon + \epsilon\mathbb{B}_1)\\ &= \mathrm{E}\big[e^{-Uh_\epsilon - \frac{1}{2}\|h_\epsilon\|_{\mathbb{H}}^2}\mathbb{1}\{W \in w - h_\epsilon + \epsilon\mathbb{B}_1\}\big].\end{aligned}$$

Since $Uh_\epsilon \geq 0$ on the event $W \in w - h_\epsilon + \epsilon\mathbb{B}_1$, the exponential is bounded above by $e^{-\frac{1}{2}\|h_\epsilon\|_{\mathbb{H}}^2}$. Furthermore, the probability $\mathrm{E}[\mathbb{1}\{W \in w - h_\epsilon + \epsilon\mathbb{B}_1\}]$ is bounded above by $\mathrm{P}(W \in \epsilon\mathbb{B}_1)$, by Anderson's lemma. ☐

By Lemma I.27 a ball of radius ϵ around a point h in the unit ball $\mathbb{H}_1$ of the RKHS contains mass at least $e^{-1/2}\mathrm{P}(\|W\| < \epsilon) = e^{-1/2}e^{-\varphi_0(\epsilon)}$. One can place $D(2\epsilon, \mathbb{H}_1, \|\cdot\|)$ points in $\mathbb{H}_1$ so that their surrounding balls of radius ϵ are disjoint. The law of total probability then gives that $1 \geq D(2\epsilon, \mathbb{H}_1, \|\cdot\|)e^{-1/2}e^{-\varphi_0(\epsilon)}$, or $\varphi_0(\epsilon) \geq \log D(2\epsilon, \mathbb{H}_1, \|\cdot\|) - 1/2$. The following two results refine this estimate and also give a bound in the other direction. The first lemma roughly shows that

$$\varphi_0(2\epsilon) \lesssim \log N\Big(\frac{\epsilon}{\sqrt{2\varphi_0(\epsilon)}}, \mathbb{H}_1, \|\cdot\|\Big) \lesssim \varphi_0(\epsilon),$$

but this is true only if the modulus and entropy are sufficiently regular functions.

Lemma I.29 *Let $f\colon (0,\infty) \to (0,\infty)$ be regularly varying*[1] *at zero. Then*

(i) $\log N(\epsilon/\sqrt{2\varphi_0(\epsilon)}, \mathbb{H}_1, \|\cdot\|) \gtrsim \varphi_0(2\epsilon)$;
(ii) *if* $\varphi_0(\epsilon) \lesssim f(\epsilon)$, *then* $\log N(\epsilon/\sqrt{f(\epsilon)}, \mathbb{H}_1, \|\cdot\|) \lesssim f(\epsilon)$;
(iii) *if* $\log N(\epsilon, \mathbb{H}_1, \|\cdot\|) \gtrsim f(\epsilon)$, *then* $\varphi_0(\epsilon) \gtrsim f(\epsilon/\sqrt{\varphi_0(\epsilon)})$;
(iv) *if* $\log N(\epsilon, \mathbb{H}_1, \|\cdot\|) \lesssim f(\epsilon)$, *then* $\varphi_0(2\epsilon) \lesssim f(\epsilon/\sqrt{\varphi_0(\epsilon)})$.

Lemma I.30 *For $\alpha > 0$ and $\beta \in \mathbb{R}$, as $\epsilon \downarrow 0$, $\varphi_0(\epsilon) \asymp \epsilon^{-\alpha}(\log_-\epsilon)^\beta$ if and only if* $\log N(\epsilon, \mathbb{H}_1, \|\cdot\|) \asymp \epsilon^{-2\alpha/(2+\alpha)}(\log_-\epsilon)^{2\beta/(2+\alpha)}$.

Proofs See Kuelbs and Li (1993) and Li and Linde (1998). ☐

The following result shows that the concentration function of the sum of independent Gaussian random elements can be estimated from the concentration functions of the components.

Lemma I.31 *Let $W = \sum_{i=1}^N W_i$ be the sum of finitely many independent Gaussian random elements in a separable Banach space $(\mathbb{B}, \|\cdot\|)$ with concentration functions φ_{i,w_i} for given $w_i in\mathbb{B}$. Then, the concentration function φ_w of W around $w = \sum_{i=1}^N w_i$ satisfies*

$$\varphi_w(N\epsilon) \leq \sum_{i=1}^N \varphi_{i,w_i}(\epsilon/2).$$

Proof By the independence of the W_is, $\mathrm{P}(\|W - w\| < N\epsilon) \geq \prod_{i=1}^N \mathrm{P}(\|W_i - w_i\| < \epsilon)$. Next two applications of Lemma I.28 lead to the desired bound. ☐

[1] A function $f(x)$ is called regularly varying at zero if $\lim_{t\to 0} f(tx)/f(t) = x^\alpha$ for some $\alpha > 0$. Examples include polynomial functions.

I.8 Historical Notes

The main part of this appendix is based on van der Vaart and van Zanten (2008b), which reviews literature on reproducing kernel Hilbert spaces. Overviews of Gaussian process theory are given by Li and Shao (2001) and Lifshits (2012). Key original references include Kuelbs and Li (1993); Ledoux and Talagrand (1991); Borell (1975); Li and Linde (1998); Kuelbs et al. (1994); Kuelbs and Li (1993). The proof of Proposition I.20 is taken from Proposition 2.1 in de Acosta (1983). For a version of the proposition for processes, see van der Vaart and van Zanten (2008b), Lemma 3.1. Lemma I.26 is taken from Castillo (2008).

Problems

I.1 In Example I.14, show that the square root $S^{1/2}$ of the operator S can be described as having the same eigenfunctions as S with eigenvalues the square roots of the eigenvalues of S.

I.2 (van der Vaart and van Zanten 2008b) Show that the ℓ_2-linear independence assumption in Theorem I.23 cannot be dropped. [Hint: If $\{h_j\}$ is not ℓ_2-linearly independent, then the RKHS is $\overline{\text{lin}}\{\sum_{i=1}^{\infty} w_i h_i : w \in \overline{\text{lin}}\{b^*(h_i) : b^* \in \mathbb{B}^*\}\}$ with squared norm $\sum_{i=1}^{\infty} w_i^2$. Taking these linear combinations for all $w \in \ell_2$ gives the same set, but the ℓ_2-norm should be computed for a projected w.]