Constr Approx (2012) 35:363–417
DOI 10.1007/s00365-012-9153-3

CONSTRUCTIVE
APPROXIMATION

Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs

Ingo Steinwart · Clint Scovel

Received: 29 October 2010 / Revised: 13 October 2011 / Accepted: 15 November 2011 /
Published online: 16 February 2012

Abstract Given a compact metric space X and a strictly positive Borel measure ν on X, Mercer's classical theorem states that the spectral decomposition of a positive self-adjoint integral operator $T_k : L_2(\nu) \to L_2(\nu)$ of a continuous k yields a series representation of k in terms of the eigenvalues and -functions of T_k. An immediate consequence of this representation is that k is a (reproducing) kernel and that its reproducing kernel Hilbert space can also be described by these eigenvalues and -functions. It is well known that Mercer's theorem has found important applications in various branches of mathematics, including probability theory and statistics. In particular, for some applications in the latter areas, however, it would be highly convenient to have a form of Mercer's theorem for more general spaces X and kernels k. Unfortunately, all extensions of Mercer's theorem in this direction either stick too closely to the original topological structure of X and k, or replace the absolute and uniform convergence by weaker notions of convergence that are not strong enough for many statistical applications. In this work, we fill this gap by establishing several Mercer type series representations for k that, on the one hand, make only very mild assumptions on X and k, and, on the other hand, provide convergence results that are strong enough for interesting applications in, e.g., statistical learning theory. To illustrate the latter, we first use these series representations to describe ranges of fractional powers of T_k in terms of interpolation spaces and investigate under which conditions these interpolation spaces are contained in $L_\infty(\nu)$.

Communicated by G. Kerkyacharian.

I. Steinwart (✉)
Institut für Stochastik und Anwendungen, Fakultät für Mathematik und Physik, Universität Stuttgart,
Pfaffenwaldring 57, 70569 Stuttgart, Germany
e-mail: ingo.steinwart@mathematik.uni-stuttgart.de

C. Scovel
Information Sciences Group CCS-3, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
e-mail: jcs@lanl.gov

For these two results, we then discuss applications related to the analysis of so-called least squares support vector machines, which are a state-of-the-art learning algorithm. Besides these results, we further use the obtained Mercer representations to show that every self-adjoint nuclear operator $L_2(\nu) \to L_2(\nu)$ is an integral operator whose representing function k is the difference of two (reproducing) kernels.

Keywords Reproducing Kernel Hilbert spaces · Integral operators · Interpolation spaces · Eigenvalues · Statistical learning theory

Mathematics Subject Classification (2000) 46E22 · 47B32 · 47B34 · 47G10 · 46B70 · 47A70 · 68T05

1 Introduction

Reproducing kernel Hilbert spaces (RKHSs) combine two very amenable mathematical properties, namely, the geometry of Hilbert spaces and the structure of function spaces. Not surprisingly, RKHSs have thus found applications in various branches of mathematics as already pointed out by Aronszajn [4] in his seminal paper. Since then, the list has become even longer, see, e.g., the books [1–3, 21, 22, 26, 28–30, 44] and the references therein. One particular area where RKHSs have various applications is probability and statistics [7, 43], including statistical learning theory [10, 14, 36]. In view of the latter, it seems natural to investigate the relationship between RKHSs, certain spaces of integrable functions such as $L_2(\nu)$ and $L_\infty(\nu)$, and the integral operators induced by the corresponding kernels and measures. From a broad view, this is the goal of the present work.

To illustrate the situation, let us assume that we have a compact metric space (X,d), a strictly positive and finite Borel measure ν on X, and an RKHS H on X with continuous kernel $k : X \times X \to \mathbb{R}$. Then k is bounded, and hence the integral operator $T_k : L_2(\nu) \to L_2(\nu)$ defined by

$$T_k f := \int_X k(\,\cdot\,, x') f(x') \, d\nu(x'), \qquad f \in L_2(\nu), \tag{1}$$

is continuous and compact. Moreover, one can show that T_k actually maps continuously into the space $C(X)$ of continuous functions, and the symmetry of k implies that T_k is self-adjoint. In addition, it turns out that T_k is also positive; that is, $\langle T_k f, f\rangle_{L_2(\nu)} \geq 0$ for all $f \in L_2(\nu)$. The spectral theorem for self-adjoint compact operators then shows that there exists an at most countable family $(e_i)_{i\in I}$ of functions $e_i : X \to \mathbb{R}$, for which the ν-equivalence classes $[e_i]_\sim := \{f : X \to \mathbb{R} \mid \nu(\{f \neq e_i\}) = 0\}$, $i \in I$, form an orthonormal system (ONS) $([e_i]_\sim)_{i\in I}$ in $L_2(\nu)$ such that

$$T_k f = \sum_{i\in I} \mu_i \langle f, [e_i]_\sim\rangle_{L_2(\nu)} [e_i]_\sim, \qquad f \in L_2(\nu). \tag{2}$$

Here $(\mu_i)_{i\in I}$ is the family of nonzero eigenvalues (with geometric multiplicities) of T_k, and $[e_i]_\sim$ is an eigenvector to the eigenvalue μ_i for every $I \in I$. Furthermore, we have $\mu_i > 0$ by the positivity of T_k, and since T_k maps into $C(X)$, we may actually choose *continuous* functions as the representatives of the eigenvectors; that is, we may assume without loss of generality that $e_i \in C(X)$ for all $i \in I$. Moreover, Mercer's theorem, see, e.g., [25], shows that the kernel k enjoys a representation in terms of the eigenvalues and the continuous representatives $(e_i)_{i\in I}$ of the eigenvectors; namely,

$$k(x, x') = \sum_{i\in I} \mu_i e_i(x) e_i(x'), \qquad x, x' \in X, \tag{3}$$

where the convergence is *absolute and uniform*. Finally, it turns out that T_k is actually nuclear; that is, $\sum_{i\in I} \mu_i < \infty$.

In the following, we call a series representation (3) a *Mercer representation* of k. It is well known that in many cases a Mercer representation of k has far-reaching consequences. For example, one can show, see, e.g., [13, 14], that the RKHS H of k is given by

$$H = \left\{ \sum_{i\in I} a_i \mu_i^{1/2} e_i : (a_i) \in \ell_2(I) \right\} \tag{4}$$

and that the corresponding norm is

$$\left\| \sum_{i\in I} a_i \mu_i^{1/2} e_i \right\|_H = \left\| (a_i) \right\|_{\ell_2(I)},$$

where $\ell_2(I)$ denotes the Hilbert space of all 2-summable real-valued families $(a_i)_{i\in I}$. In particular, the system $(\sqrt{\mu_i} e_i)_{i\in I}$ forms an orthonormal basis (ONB) of H, and hence $[e_i]_\sim \mapsto \sqrt{\mu_i} e_i$ induces a linear isometry between a subspace of $L_2(\nu)$ and H. Furthermore, it is easy to check that this isometry equals the square root $T_k^{1/2}$ of the integral operator T_k. What happens with the images of other fractional powers T_k^β of T_k? It turns out, see [14, 34], that (4) can be used to show that these images can be closely, though not exactly, described by the real interpolation spaces $[L_2(\nu), [H]_\sim]_{p,\infty}$, where $[H]_\sim$ denotes the RKHS H when considered as a subset of $L_2(\nu)$ and p is suitably chosen. In turn, these interpolation spaces describe classes of functions that can be approximated by balls in H with a certain, controlled accuracy, see again [14, 34]. Consequently, Mercer's theorem provides a link between the spectral properties of T_k and the approximation properties of H, and this link has some far-reaching consequences for the analysis of some state-of-the-art learning algorithms, see, e.g., [9, 13, 15, 20, 35, 37] and the references therein. Besides this particular example, the Mercer representation (3) has several further applications in probability and statistics, such as the Karhunen-Loève representation for stochastic processes and its applications [7], and dimensionality reduction methods such as kernel PCA [32, 33]. While in some of these areas, X is simply a bounded and closed interval, in some other areas, even the assumption that X is a compact metric space is too restrictive. In particular, this is true for the analysis of the above mentioned learning algorithms, where one motivation for the use of kernels is that they make it possible to consider almost arbitrary sets X such as sets of graphs, strings, functions, or probability measures, see [12, 31] and the references therein. This raises the questions of whether general kernels still enjoy a series representation of the form (3) and whether the consequences of such a representation still hold in this general setting. Finding answers to these two questions is one of the goals of this paper.

Before we briefly describe the main results of this work in a somewhat informal way, let us briefly recall what is known so far. Recently, [40] extended Mercer's theorem to σ-compact metric spaces with σ-finite, strictly positive Borel measures and continuous kernels satisfying some natural integrability conditions. For example, this result includes the important case of Gaussian RBF kernels on $\mathbb{R}^d$, if the measure is actually finite. However, even with this result, the case of general *sets* X remains unclear. To have a closer look at this general situation, let us now assume that X is a measurable space[1] and that ν is a measure on X. Moreover, let H be an RKHS with measurable kernel $k : X \times X \to \mathbb{R}$ such that

$$\|k\|_{L_2(\nu\otimes\nu)} = \left(\int_X \int_X \left|k\left(x, x'\right)\right|^2 d\nu(x)\, d\nu\left(x'\right)\right)^{1/2} < \infty. \tag{5}$$

Then, the integral operator $T_k : L_2(\nu) \to L_2(\nu)$ defined by (1) is continuous, self-adjoint, and Hilbert-Schmidt, $T_k \in \mathsf{HS}(L_2(\nu))$, i.e., its eigenvalues are 2-summable, and the map $L_2(\nu \otimes \nu) \to \mathsf{HS}(L_2(\nu))$ defined by $k \mapsto T_k$ is an isometric isomorphism, see [45, Satz VI.6.3]. Consequently, the spectral theorem yields the spectral representation (2) of T_k. Moreover, one can show that k enjoys a representation of the form

$$k = \sum_{i\in I} \mu_i e_i \otimes e_i, \tag{6}$$

where the convergence is in the $L_2(\nu \otimes \nu)$-sense and $(e_i \otimes e_i)(x, x') := e_i(x)e_i(x')$. In fact, the latter representation does not only hold for kernels but for all symmetric, measurable functions $k : X \times X \to \mathbb{R}$ satisfying (5), see, e.g., [45, Chap. VI.4]. Unfortunately, however, $L_2(\nu \otimes \nu)$-convergence does not imply $\nu \otimes \nu$-almost sure pointwise convergence, and hence constructions such as (4) are, in general, impossible. To address this issue, one could be tempted to restrict considerations to bounded kernels k and finite measures ν, since in this case [19, p. 145] shows that (6) actually converges in $L_\infty(\nu\otimes\nu)$. Although $L_\infty(\nu\otimes\nu)$-convergence implies $\nu\otimes\nu$-almost sure pointwise convergence, it will turn out later in this work that even this convergence is, in general, not enough. For example, having $L_\infty(\nu \otimes \nu)$-convergence only, we can, in general, not use (μ_i) and (e_i) to reconstruct the reproducing kernel Hilbert space H nor its image $[H]_\sim$ in $L_2(\nu)$. This impossibility has some very practical consequences. Indeed, we will see around Corollary 3.8 that the right-hand side of (6) can only be used to compute the Gram matrix used in the learning algorithms mentioned above, if we have a more restricted notion of almost sure convergence.

Let us now provide a brief and informal description of our main results. To this end, we assume that the RKHS H is compactly embedded into $L_2(\nu)$; that is, the linear map $I_k : H \to L_2(\nu)$ defined by $I_k f := [f]_\sim$ is well defined and compact.

[1] We usually omit a symbol for the corresponding σ-algebra, since, in general, we do not use it.

Here, we note that this is always satisfied for bounded kernels and finite measures, but these are not the only examples of compactly embedded RKHSs. It turns out that these assumptions ensure that the integral operator $T_k : L_2(\nu) \to L_2(\nu)$ is defined, self-adjoint, compact, and positive. We thus have the spectral representation (2), and we will see that we can choose representatives e_i with $e_i \in H$ for all $i \in I$. Furthermore, the family $(\sqrt{\mu_i} e_i)$ is actually an ONS in H. In other words, we have two ONSs described by the family $(e_i)_{i\in I}$ of representatives, namely the just mentioned ONS in H and $([e_i]_\sim)_{i\in I}$ in $L_2(\nu)$. In general, neither of these ONSs are ONBs; that is, H and $L_2(\nu)$ only "share" certain subspaces with each other, and we will see that the "size" of the orthogonal complements of these subspaces will play a crucial role in our investigations. With these rather introductory remarks, we can now describe the following main results:

- The kernel k enjoys a Mercer representation (3) with pointwise absolute convergence if and only if the embedding $I_k : H \to L_2(\nu)$ is injective. Moreover, this is exactly the case when the family $(\sqrt{\mu_i} e_i)$ is actually an ONB of H.
- We always have a Mercer representation (3) with $\nu \otimes \nu$-almost sure absolute convergence, but, in general, we *cannot* ensure pointwise convergence on a set of the form $(X \setminus N) \times (X \setminus N)$, where $\nu(N) = 0$. However, for *separable* H, we actually have pointwise convergence on such a set, and the same is true for continuous kernels if the measure ν has a largest open zero set. In this regard, recall that in a rigorous statistical analysis of the learning algorithms mentioned above, one usually has to assume the separability of H anyway, see [36] for some reasons, including certain measurability aspects. In these cases, we thus obtain the Mercer representation on rectangular sets of full measure for free.
- We have already mentioned that for every Hilbert-Schmidt operator $T : L_2(\nu) \to L_2(\nu)$, there exists a $k \in L_2(\nu \otimes \nu)$ with $T = T_k$. We will show that for *self-adjoint* nuclear operators T, k can be chosen to be a function of the form $k = k^+ - k^-$, where $k^\pm$ are *kernels* with separable RKHSs that are compactly embedded into $L_2(\nu)$.
- For every RKHS, the right-hand side of (3) converges pointwise, and hence defines a new kernel k_ν with separable RKHS H_ν. The spaces H_ν and H coincide when considered as subspaces of $L_2(\nu)$; that is, $[H_\nu]_\sim = [H]_\sim$, and this relation can be used to show that the image of the fractional power $T_k^{\beta/2}$, $\beta \in (0, 1]$, equals the real interpolation space $[L_2(\nu), [H]_\sim]_{\beta,2}$.
- If the eigenvalues $(\mu_i)_{i\in I}$ are β-summable, that is, $\sum_{i\in I} \mu_i^\beta < \infty$, then there exists a separable RKHS $\bar{H}_\nu^\beta$ with $[\bar{H}_\nu^\beta]_\sim = [L_2(\nu), [H]_\sim]_{\beta,2}$, and its kernel $\bar{k}_\nu^\beta$ represents T_k^β; that is, $T_k^\beta = T_{\bar{k}_\nu^\beta}$. Furthermore, this kernel enjoys a pointwise absolute convergent Mercer representation

  $$\bar{k}_\nu^\beta(x, x') = \sum_{i\in I} \mu_i^\beta \bar{e}_i(x)\bar{e}_i(x'), \quad x, x' \in X,$$

  where $\bar{e}_i : X \to \mathbb{R}$ are suitable functions with $[\bar{e}_i]_\sim = [e_i]_\sim$ for all $i \in I$.
- Suppose that ν is a σ-finite measure on a complete σ-algebra $\mathcal{A}$ on X. Then the space $[L_2(\nu), [H]_\sim]_{\beta,2}$ is continuously embedded into $L_\infty(\nu)$ if and only if there

exists such a kernel $\bar{k}_\nu^\beta$ and this kernel is ν-almost surely bounded on the diagonal $x = x'$. Moreover, in this case the eigenvalues $(\mu_i)_{i\in I}$ are β-summable.
- As an application, we finally show that if $[L_2(\nu), [H]_\sim]_{\beta,2}$ is continuously embedded into $L_\infty(\nu)$, then infinite-sample solutions of certain learning algorithms, namely least-squares support vector machines (LS-SVMs), have smaller supremum norms than previously anticipated. To illustrate this improvement, we briefly discuss some implications for the statistical analysis of these learning methods.

This list indicates that ν-zero sets play an important role in our analysis. One of the reasons for this is that the orthogonal complement of the space spanned by the ONS $(\sqrt{\mu_i}e_i)$ consists exactly of those $f \in H$ that satisfies $[f]_\sim = 0$. As a consequence, we will meticulously distinguish between functions and ν-equivalence classes of functions.

The rest of this work is organized as follows: in Sect. 2, we will introduce basic concepts and results related to kernels, integral operators, and ONSs in $L_2(\nu)$. Then, in Sect. 3, we will present several results related to Mercer representations and the space H_ν. There, we will also present the integral representation of self-adjoint nuclear operators. In Sect. 4, we will investigate the spaces associated with fractional powers of T_k and the existence of the kernels $\bar{k}_\nu^\beta$. The boundedness of these kernels are then discussed in Sect. 5. Finally, all proofs can be found in Sect. 6, and some auxiliary material on the spectral theorem and on liftings of $L_\infty(\nu)$ are compiled in two appendices.

2 Compactly Embedded RKHSs

In this section, we investigate RKHSs that are compactly embedded into $L_2(\nu)$. In particular, we will define the corresponding integral operator and present some basic properties of this operator. We will also introduce some classes of RKHSs that are compactly embedded into $L_2(\nu)$.

Throughout this work, we assume that $(X, \mathcal{A})$ is a measurable space and ν is a measure on $(X, \mathcal{A})$. Since in most situations, the specific form of the σ-algebra $\mathcal{A}$ does not matter, we usually omit $\mathcal{A}$ and simply speak of a measurable space X and a measure ν on X. Moreover, if we have a topology τ on X, then we always assume that $\mathcal{A}$ is the associated Borel σ-algebra. We sometimes also need complete σ-algebras. Therefore, let us recall that a σ-algebra $\mathcal{A}$ is ν-complete, if, for every $A \subset X$ for which there exists an $N \in \mathcal{A}$ such that $A \subset N$ and $\nu(N) = 0$, we have $A \in \mathcal{A}$. Furthermore, we denote by $\mathcal{L}_2(\nu)$ the set of all measurable functions $f : X \to \mathbb{R}$ with $\int |f|^2 \, d\nu < \infty$. For $f \in \mathcal{L}_2(\nu)$, we further write

$$[f]_\sim := \big\{ g \in \mathcal{L}_2(\nu) : \nu\big(\{f \neq g\}\big) = 0 \big\}$$

for the ν-equivalence class of f. As usual, we write $L_2(\nu) := \mathcal{L}_2(\nu)_{/\sim}$ for the corresponding quotient space and $\|\cdot\|_{\mathcal{L}_2(\nu)}$ and $\|\cdot\|_{L_2(\nu)}$ for the (semi-)norm of $\mathcal{L}_2(\nu)$ and $L_2(\nu)$, respectively. Finally, we define the ν-equivalence classes in other spaces of measurable functions analogously, where we note, however, that in some cases using the symbol $[f]_\sim$ for such classes is actually a slight abuse of notation.

In the following, we say that a Banach space F is continuously embedded into a Banach space E if $F \subset E$ and the identity map $\mathrm{id} : F \to E$ is continuous. In this case, we write $F \hookrightarrow E$.

Recall that Hilbert function spaces H over X are Hilbert spaces that consists of functions $f : X \to \mathbb{R}$. Moreover, a Hilbert function space H over X is an *RKHS* if and only if all evaluation functionals $\delta_x : H \to \mathbb{R}$, $x \in X$, defined by $\delta_x(f) := f(x)$, $f \in H$, are continuous. In addition, a map $k : X \times X \to \mathbb{R}$ is called a *kernel* if there exists a Hilbert space $\tilde{H}$ and a map $\tilde{\Phi} : X \to \tilde{H}$ such that

$$k\big(x, x'\big) = \big\langle \tilde{\Phi}(x), \tilde{\Phi}\big(x'\big)\big\rangle_{\tilde{H}}, \quad x, x' \in X.$$

In the statistical learning theory literature, $\tilde{H}$ is often called *feature space* of k and $\tilde{\Phi}$ *feature map* of k. It is well known that for every RKHS, there exists exactly one kernel k on X such that $k(\cdot, x) \in H$ and

$$f(x) = \big\langle f, k(\cdot, x)\big\rangle_H, \quad f \in H, x \in X.$$

Conversely, for every kernel on X there exists exactly one RKHS H over X for which the above equation holds. In other words, there is a one-to-one relation between kernels and RKHSs. Fortunately, there exist various textbooks that discuss basic (and advanced) properties of RKHSs and their kernels, see, e.g., the books mentioned in the introduction. We thus omit a more detailed discussion on this topic and assume that the reader is familiar with these objects. Finally, we mention that we will mainly follow the RKHS notation and exposition of [36].

Let us now introduce the class of RKHSs with which we will mainly deal.

Definition 2.1 Let X be a measurable space, ν be a measure on X, and H be an RKHS over X with measurable kernel $k : X \times X \to \mathbb{R}$. We say that H is embedded into $L_2(\nu)$ if $[f]_\sim \in L_2(\nu)$ for all $f \in H$ and the linear operator

$$\begin{aligned} I_k : H &\to L_2(\nu), \\ f &\mapsto [f]_\sim \end{aligned}$$

is continuous. We write $[H]_\sim$ for its image; that is, $[H]_\sim := \{[f]_\sim : f \in H\}$. Moreover, we say that H is compactly embedded into $L_2(\nu)$ if I_k is compact.

Recall that one usually calls the identity map $\mathrm{id} : E \to F$ between two spaces $E \subset F$ an embedding if this map is continuous. Intuitively, Definition 2.1 does the same, but formally it does not, since H consists of functions and $L_2(\nu)$ does not. We will see throughout this paper that this subtle difference has far-reaching consequences.

If H is embedded into $L_2(\nu)$, we have $[k(x, \cdot)]_\sim \in L_2(\nu)$ for all $x \in X$, and hence Hölder's inequality shows that $[k(x, \cdot)]_\sim f(\cdot) \in L_1(\nu)$ for all $x \in X$ and all $f \in L_2(\nu)$. Consequently, $T_k f$ in (1) is defined for all $f \in L_2(\nu)$. The following fundamental lemma, which is essentially taken from [16], shows that $T_k : L_2(\nu) \to L_2(\nu)$ is actually continuous, self-adjoint, and positive.

Lemma 2.2 (Properties of the integral operator) *Let X be a measurable space, ν be a measure on X, and H be an RKHS over X with measurable kernel $k : X \times X \to \mathbb{R}$. Assume that H is embedded into $L_2(\nu)$. Then* (1) *defines a continuous, positive, and self-adjoint operator $T_k : L_2(\nu) \to L_2(\nu)$. Moreover, the adjoint $S_k := I_k^* : L_2(\nu) \to H$ of the embedding $I_k : H \to L_2(\nu)$ satisfies*

$$S_k f(x) = \int_X k\big(x, x'\big) f\big(x'\big)\, d\nu\big(x'\big), \qquad f \in L_2(\nu), x \in X, \tag{7}$$

and thus we have the following factorization:

$$\begin{tikzcd}
L_2(\nu) \arrow[rr, "T_k"] \arrow[dr, "S_k"'] & & L_2(\nu) \\
& H \arrow[ur, "S_k^*"'] &
\end{tikzcd}$$

Consequently, if H is compactly embedded into $L_2(\nu)$, then T_k is compact.

Let us now discuss some cases in which an RKHS is compactly embedded into some $L_2(\nu)$. The first sufficient condition, which is a straightforward adaptation from [16], is usually easy to check.

Lemma 2.3 (Integrability on the diagonal) *Let X be a measurable space, ν be a measure on X, and H be an RKHS over X with measurable kernel $k : X \times X \to \mathbb{R}$. If*

$$\|k\|_{\mathcal{L}_2(\nu)} := \left(\int_X k(x, x)\, d\nu(x) \right)^{1/2} < \infty,$$

then H is compactly embedded into $L_2(\nu)$. Moreover, I_k and S_k are Hilbert-Schmidt, and T_k is nuclear. Finally, we have $\|k\|_{\mathcal{L}_2(\nu\otimes\nu)} \le \|k\|^2_{\mathcal{L}_2(\nu)}$.

Note that if the kernel k is measurable and bounded, that is,

$$\|k\|_\infty := \sup_{x \in X} \sqrt{k(x, x)} < \infty,$$

then, for all finite measures ν on X, we have $\|k\|_{\mathcal{L}_2(\nu)} \le \sqrt{\nu(X)} \cdot \|k\|_\infty$, and hence its RKHS H is compactly embedded into $L_2(\nu)$. Moreover, there exist kernels with $\|k\|_{\mathcal{L}_2(\nu)} < \infty$ and $\|k\|_\infty = \infty$, and hence the notion of compactly embedded RKHSs is strictly weaker than the boundedness of kernels. Below, we will further see that not every compactly embedded RKHS satisfies $\|k\|_{\mathcal{L}_2(\nu)} < \infty$.

In order to present a second sufficient condition for compactly embedded RKHSs, we say that an at most countable family $(\alpha_i)_{i\in I} \subset (0, \infty)$ converges to 0 if either $I = \{1, \dots, n\}$ or $I = \mathbb{N} := \{1, 2, \dots\}$ and $\lim_{i\to\infty} \alpha_i = 0$. Analogously, when we consider an at most countable family $(e_i)_{i\in I}$, we always assume without loss of generality that either $I = \{1,\dots,n\}$ or $I = \mathbb{N}$. Here we recall that the convergence of a family $(\alpha_i)_{i\in I} \subset (0,\infty)$ does not change under permutations of its index set I, and hence these assumptions are consistent. With these preparations, the second sufficient condition, which is of less practical interest but important for some of our constructions, reads as follows:

Lemma 2.4 *Let X be a measurable space, ν be a measure on X, and H be an RKHS over X with measurable kernel $k : X \times X \to \mathbb{R}$. Assume that there exists a (not necessarily countable) ONB $(e_j)_{j\in J}$ of H such that*

$$\big\langle [e_i]_\sim, [e_j]_\sim \big\rangle_{L_2(\nu)} = 0 \quad \textit{if } i \neq j.$$

Then the following statements are true:

(i) *If $\sup_{j\in J} \|[e_j]_\sim\|_{L_2(\nu)} < \infty$, then S_k^* is continuous; that is, H is embedded into $L_2(\nu)$.*
(ii) *If J is countable, that is, H is separable, and the family $(\|[e_j]_\sim\|_{L_2(\nu)})_{j\in J}$ converges to 0, then H is compactly embedded into $L_2(\nu)$.*

Our next goal is to construct RKHSs that are compactly embedded into $L_2(\nu)$. To this end, we write, for an arbitrary index set J and a real-valued family $a := (a_j)_{j\in J}$,

$$\|a\|_{\ell_2(J)} := \left(\sum_{j\in J} |a_j|^2\right)^{1/2}.$$

As usual, we further denote the Hilbert space of all 2-summable real-valued families $(a_j)_{j\in J}$ on J by $\ell_2(J)$; that is, $\ell_2(J) := \{(a_j)_{j\in J} \subset \mathbb{R} : \|(a_j)_{j\in J}\|_{\ell_2(J)} < \infty\}$. Clearly, $\ell_2(J) = \mathcal{L}_2(\nu) = L_2(\nu)$, where ν is the counting measure on J, and it is easy to check that $\ell_2(J)$ is an RKHS, which is separable if and only if J is at most countable. Moreover, for a bounded family $\mu := (\mu_j)_{j\in J} \subset [0,\infty)$ and $b := (b_j) := (b_j)_{j\in J} \subset \mathbb{R}$, we define

$$\|b\|^2_{\ell_2(\mu^{-1})} := \sum_{j\in J} \frac{b_j^2}{\mu_j},$$

where we use the rather common convention $0/0 := 0$. Note that for $(b_j)_{j\in J} \in \ell_2(\mu^{-1})$, this convention implies $b_j = 0$ for all $j \in J$ for which $\mu_j = 0$. Moreover, it is easy to see that $\|\cdot\|_{\ell_2(\mu^{-1})}$ is a Hilbert space norm on the set

$$\ell_2\big(\mu^{-1}\big) := \big\{(b_j)_{j\in J} : \big\|(b_j)\big\|_{\ell_2(\mu^{-1})} < \infty\big\}.$$

Finally, we have $\mu_{j_0} \le \|\mu\|_\infty := \sup_{j\in J} \mu_j$ for all $j_0 \in J$, and from this it is easy to conclude that both $\ell_2(\mu^{-1}) \subset \ell_2(J)$ and

$$\|b\|_{\ell_2(J)} \le \|\mu\|_\infty^{1/2} \|b\|_{\ell_2(\mu^{-1})}, \quad b \in \ell_2\big(\mu^{-1}\big).$$

For the construction of compactly embedded RKHSs, we also need the following simple but powerful lemma:

Lemma 2.5 *Let X be a nonempty set and $(f_i)_{i\in I}$ be an at most countable family of functions mapping from X to $\mathbb{R}$. Then the following statements are equivalent*:

(i) *The series $\sum_{i\in I} a_i f_i(x)$ converges for all $(a_i)\in \ell_2(I)$ and all $x\in X$.*
(ii) *For all $x\in X$, we have $\sum_{i\in I} f_i^2(x)<\infty$.*
(iii) *The series $\sum_{i\in I} f_i(x)f_i(x')$ converges for all $x,x'\in X$.*

Moreover, if one (and thus all) of the statements above is true, all series above converge absolutely.

With the help of these preparations, we will now show how at most countable ONSs in $L_2(\nu)$ can be used to define compactly embedded, separable RKHSs that "share" a subspace with $L_2(\nu)$.

Lemma 2.6 *Let X be a measurable space, ν be a measure on X, and $(e_i)_{i\in I}$ be an at most countable family of functions such that $([e_i]_\sim)_{i\in I}$ is an ONS in $L_2(\nu)$. Furthermore, let $(\mu_i)_{i\in I}\subset(0,\infty)$ be a family converging to 0 such that*

$$\sum_{i\in I}\mu_i e_i^2(x)<\infty,\qquad x\in X. \tag{8}$$

Then the space

$$H:=\left\{\sum_{i\in I} a_i\mu_i^{1/2}e_i : (a_i)\in\ell_2(I)\right\}=\left\{\sum_{i\in I} b_i e_i : (b_i)\in\ell_2\big(\mu^{-1}\big)\right\} \tag{9}$$

equipped with the norm

$$\left\|\sum_{i\in I} b_i e_i\right\|_H := \big\|(b_i)\big\|_{\ell_2(\mu^{-1})} \tag{10}$$

is a separable RKHS that is compactly embedded into $L_2(\nu)$. Moreover, $(\sqrt{\mu_i}e_i)$ is an ONB of H and the (measurable) kernel k of H is given by the pointwise convergent series representation

$$k\big(x,x'\big)=\sum_{i\in I}\mu_i e_i(x)e_i\big(x'\big),\qquad x,x'\in X. \tag{11}$$

Definition 2.7 (RKHS induced by an ONS) Let X be a measurable space, ν be a measure on X, and $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I}\subset(0,\infty)$ be as in Lemma 2.6. Then H defined by (9) and (10) is called the RKHS that is induced by $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I}$. Analogously, k defined by (11) is called the kernel induced by $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I}$.

Remark 2.8 Actually, the construction of Lemma 2.6 can be made under much weaker assumptions. Indeed, if $(f_j)_{j\in J}$ is a (not necessarily countable) family of functions $f_j : X \to \mathbb{R}$ with $(f_j(x))_{j\in J} \in \ell_2(J)$ for all $x \in X$, then $k(x,x') := \sum_{j\in J} f_j(x) f_j(x')$ defines a kernel on X. Furthermore, the space $\ell_2(J)$ is a feature space of k with feature map given by $\tilde{\Phi}(x) = (f_j(x))_{j\in J}$. Together with [36, Theorem 4.21], this can be used to show that the set of functions

$$H := \left\{ \sum_{j\in J} a_j f_j : (a_j) \in \ell_2(J) \right\}$$

is the RKHS of k. In general, however, the representation of an $f \in H$ by the family $(f_j)_{j\in J}$ is *not unique*, and hence the norm of H needs to be defined by the infimum taken over all such representations. As a consequence, $(f_j)_{j\in J}$ does not need to be an ONB of H, and even if it is an ONB of H, it does, in general, not satisfy $\langle [f_i]_\sim, [f_j]_\sim \rangle_{L_2(\nu)} = 0$ for $i \neq j$.

Let us now discuss some consequences of Lemma 2.6. We begin with the following example that shows that there exists separable, compactly embedded RKHSs with $\|k\|_{\mathcal{L}_2(\nu)} = \infty$.

Example 2.9 Consider Lemma 2.6 for $X := I := \mathbb{N}$ and $e_i := \mathbf{1}_{\{i\}}$. Moreover, let ν be the counting measure on X, and $\mu_i := i^{-1}$, $i \in I$. Then $([e_i]_\sim)_{i\in I}$ is an ONB of $L_2(\nu) = \ell_2(I)$, and (8) is satisfied since $\sum_{i\in I} \mu_i e_i^2(x) = \mu_x$ for all $x \in X$. Consequently, the induced kernel satisfies $k(x,x) = \mu_x$, and since $\sum_{i\in I} \mu_i = \infty$, we conclude that $\|k\|_{\mathcal{L}_2(\nu)} = \infty$.

A first consequence of Lemma 2.6 is the following result that compares different notions of convergence in Mercer representations of kernels. The remarkable message in this result is that pointwise convergence in a Mercer representation implies the usually stronger convergence in H, if the representing set of functions defines an ONS in $L_2(\nu)$.

Corollary 2.10 (Equivalent notions of convergence in Mercer representations) *Let X be a measurable space, ν be a measure on X, and $(e_i)_{i\in I}$ be an at most countable family of functions such that $([e_i]_\sim)_{i\in I}$ is an ONS in $L_2(\nu)$. In addition, let $(\mu_i)_{i\in I} \subset (0,\infty)$ be a family converging to 0 and k be a measurable kernel on X with RKHS H. Then the following statements are equivalent*:

(i) *For all $x, x' \in X$, we have $k(x,x') = \sum_{i\in I} \mu_i e_i(x) e_i(x')$.*
(ii) *For all $x, x' \in X$, we have $k(x,x') = \sum_{i\in I} \mu_i e_i(x) e_i(x')$, where the convergence is absolute.*
(iii) *For all $x, x' \in X$, we have $k(x,x') = \sum_{i\in I} \mu_i e_i(x) e_i(x')$, and for all $x \in X$, the convergence is uniform in x' on all $A \subset X$ with $\sup_{x''\in A} k(x'',x'') < \infty$.*
(iv) *For all $x \in X$, we have $k(\cdot,x) = \sum_{i\in I} \mu_i e_i(x) e_i$, where the convergence is in H.*
(v) *For all $x \in X$, we have $k(\cdot,x) = \sum_{i\in I} \mu_i e_i(x) e_i$, where the convergence is unconditional in H.*

So far, we have only investigated induced RKHSs and their kernels. Let us now have a look at the spectral properties of the integral operators of induced kernels.

Theorem 2.11 (Integral operators of induced kernels) *Let X be a measurable space, ν be a measure on X, and $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as in Lemma 2.6. Furthermore, let H be the RKHS induced by $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I}$ and k be its kernel. Then the following statements are true*:

(i) *The operator $S_k : L_2(\nu) \to H$ can be computed by*

$$S_k f = \sum_{i\in I} \mu_i \langle f, [e_i]_\sim \rangle_{L_2(\nu)} e_i, \quad f \in L_2(\nu), \tag{12}$$

where the convergence is in H. Moreover, $(\mu_i)_{i\in I}$ is the family of nonzero eigenvalues (with geometric multiplicities) of the integral operator $T_k : L_2(\nu) \to L_2(\nu)$, and $([e_i]_\sim)_{i\in I}$ is an ONS consisting of the corresponding eigenvectors. Consequently, T_k is given by

$$T_k f = \sum_{i\in I} \mu_i \langle f, [e_i]_\sim \rangle_{L_2(\nu)} [e_i]_\sim, \quad f \in L_2(\nu), \tag{13}$$

where the convergence is in $L_2(\nu)$.

(ii) *The subspace $\overline{\mathrm{span}\{[e_i]_\sim : i \in I\}}$ of $L_2(\nu)$ is isometrically isomorphic to H via the obvious restriction of the map $S_k^{1/2} : L_2(\nu) \to H$ defined by*

$$S_k^{1/2} f := \sum_{i\in I} \mu_i^{1/2} \langle f, [e_i]_\sim \rangle_{L_2(\nu)} e_i, \quad f \in L_2(\nu),$$

where the convergence is in H.

(iii) *We have $\|k\|^2_{\mathcal{L}_2(\nu)} = \sum_{i\in I} \mu_i$, where both values may be infinite.*

Lemma 2.6 showed that every ONS $([e_i]_\sim)_{i\in I}$ in $L_2(\nu)$ together with a family $(\mu_i)_{i\in I} \subset (0,\infty)$ converging to 0 defines an RKHS that is compactly embedded into $L_2(\nu)$. Moreover, in Theorem 2.11, we have seen that in this case the operator T_k enjoys a spectral representation (13) in terms of $([e_i]_\sim)_{i\in I}$ and $(\mu_i)_{i\in I}$. On the other hand, if we start with a kernel k for which the integral operator T_k is compact, then the spectral theorem for self-adjoint compact operators on Hilbert spaces shows that T_k always enjoys such a representation in terms of its eigenvalues and an arbitrary ONS of associated eigenfunctions. The following result provides some additional properties of this representation in the case where the RKHS of k is compactly embedded into $L_2(\nu)$:

Lemma 2.12 *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X with RKHS H such that H is compactly embedded into $L_2(\nu)$. Let $(\mu_i)_{i\in I} \subset (0,\infty)$ be an at most countable family converging to 0 with $\mu_1 \ge \mu_2 \ge \cdots > 0$ and $(\tilde e_i)_{i\in I}$ be an arbitrary family of functions such that $([\tilde e_i]_\sim)_{i\in I}$ is an ONS in $L_2(\nu)$ and*

$$T_k f = \sum_{i\in I} \mu_i \langle f, [\tilde e_i]_\sim \rangle_{L_2(\nu)} [\tilde e_i]_\sim, \quad f \in L_2(\nu), \tag{14}$$

with convergence in $L_2(\nu)$. Then the following statements are true:

(i) *There always exist families* $(\tilde{e}_i)_{i\in I}$ *and* $(\mu_i)_{i\in I} \subset (0,\infty)$ *satisfying these assumptions.*
(ii) *The family* $(\mu_i)_{i\in I}$ *necessarily consists of the nonzero eigenvalues* (*with geometric multiplicities*) *of* T_k.
(iii) *The family* $([\tilde{e}_i]_\sim)_{i\in I}$ *necessarily consists of the corresponding eigenvectors of* T_k.
(iv) *There exists a family* $(e_i)_{i\in I} \subset H$ *such that* $(\sqrt{\mu_i}e_i)_{i\in I}$ *is an ONS in* H *and* $[e_i]_\sim = [\tilde{e}_i]_\sim$ *for all* $i \in I$.

In particular, (14) *also holds for the family* $([e_i]_\sim)_{i\in I}$, *and hence* $([e_i]_\sim)_{i\in I}$ *is an ONS in* $L_2(\nu)$ *that consists of eigenvectors of* T_k *corresponding to* $(\mu_i)_{i\in I} \subset (0,\infty)$. *In addition*, *we have*

$$\mu_i e_i = S_k[e_i]_\sim, \quad i \in I, \tag{15}$$

$$\ker S_k = \ker T_k, \tag{16}$$

$$\overline{\operatorname{ran} S_k} = \overline{\operatorname{span}\{\sqrt{\mu_i}e_i : i \in I\}}, \tag{17}$$

$$\overline{\operatorname{ran} S_k^*} = \overline{\operatorname{span}\big\{[e_i]_\sim : i \in I\big\}}, \tag{18}$$

$$\ker S_k^* = (\overline{\operatorname{ran} S_k})^\perp, \tag{19}$$

$$\overline{\operatorname{ran} S_k^*} = (\ker S_k)^\perp, \tag{20}$$

where the closures and orthogonal complements are taken in the spaces the objects are naturally contained in; *that is*, (17) *and* (19) *are considered in* H, *while* (18) *and* (20) *are considered in* $L_2(\nu)$.

Clearly, there are multiple ONSs $([\tilde{e}_i]_\sim)_{i\in I}$ that satisfy (14), and consequently, the family $(e_i)_{i\in I} \subset H$ is not uniquely determined. However, once we have picked an ONS $([\tilde{e}_i]_\sim)_{i\in I}$ in (14), the family $(e_i)_{i\in I} \subset H$ *is uniquely determined* since (15) shows $\mu_i e_i = S_k[e_i]_\sim = S_k[\tilde{e}_i]_\sim$ for all $i \in I$. In other words, there is only one representative e_i of $[\tilde{e}_i]_\sim$ in H, and hence the ambiguity for $(e_i)_{i\in I} \subset H$ comes solely from the ambiguity of the ONS $([\tilde{e}_i]_\sim)_{i\in I}$, that is, from the freedom of choosing an ONB in each eigenspace of T_k. In particular, the Mercer representations established in the next section are unique modulo this freedom.

3 Mercer Representations on General Domains

In this section, we investigate under which conditions we have a Mercer representation that converges pointwise, or at least pointwise on a set of the form $(X\setminus N) \times (X\setminus N)$. We will see that the first type of convergence holds if and only if the ONS $(\sqrt{\mu_i}e_i)_{i\in I}$ is an ONB of H, while the second convergence holds if H is separable or k is continuous and ν has a largest open zero set. Moreover, we will show that the series of a Mercer representation of a kernel k always converges pointwise to a kernel, which in general, however, is different from k. Despite this difference, the RKHS of the new kernel equals the original RKHS of the kernel when both spaces are considered as subspaces of $L_2(\nu)$. Finally, we will show that self-adjoint nuclear operators on $L_2(\nu)$ enjoy an integral representation in terms of differences of kernels.

Our first result characterizes the situation in which the ONSs $([e_i]_\sim)_{i\in I}$ or $(\sqrt{\mu_i}e_i)_{i\in I}$ are actually ONBs of $L_2(\nu)$ or H, respectively.

Theorem 3.1 (Pointwise convergent Mercer representation) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X with RKHS H. Assume that H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i\in I}\subset H$ and $(\mu_i)_{i\in I}\subset(0,\infty)$ be as in Lemma 2.12. Then the following statements are equivalent*:

(i) *The family $([e_i]_\sim)_{i\in I}$ is an ONB of $L_2(\nu)$.*
(ii) *The operator $S_k^*: H\to L_2(\nu)$ has a dense image in $L_2(\nu)$; that is, $[H]_\sim$ is a dense subset of $L_2(\nu)$.*

In addition, the following statements are equivalent:

(i) *The family $(\sqrt{\mu_i}e_i)_{i\in I}$ is an ONB of H.*
(ii) *The operator $S_k^*: H\to L_2(\nu)$ is injective.*
(iii) *The kernel k enjoys a pointwise convergent Mercer representation*; *that is, for all $x,x'\in X$, we have*

$$k\bigl(x,x'\bigr)=\sum_{i\in I}\mu_i e_i(x)e_i\bigl(x'\bigr). \tag{21}$$

Recall that denseness of $[H]_\sim$ in $L_2(\nu)$ is a highly desirable property in statistical learning theory, since it can be used to establish asymptotic learning results, namely consistency, for learning methods such as LS-SVMs, see, e.g., [36, Chaps. 6 & 7]. For our purposes, however, the second set of equivalences is more interesting, since it characterizes pointwise convergent Mercer representations. In this respect, recall Corollary 2.10, which showed that the pointwise convergence in (21) implies stronger notions of convergence, such as unconditional convergence in H.

In general, the map $S_k^*: H\to L_2(\nu)$ is *not* injective, and hence $(\sqrt{\mu_i}e_i)_{i\in I}$ is not an ONB of H, which, by Theorem 3.1, shows that k does not have a pointwise convergent Mercer representation (21). Our next goal is to show that k enjoys, however, a weak form of a Mercer representation. To this end, recall that every ONS can be extended to an ONB. Let us assume that we have fixed such an extension of the ONS $(\sqrt{\mu_i}e_i)_{i\in I}$ in H; i.e., we have an ONS $(\bar e_j)_{j\in J}$ in H such that the union of $(\sqrt{\mu_i}e_i)_{i\in I}$ and $(\bar e_j)_{j\in J}$ is an ONB of H. Then [36, Theorem 4.20] or Theorem 6.2 below show that

$$k(\cdot,x)=\sum_{i\in I}\mu_i e_i(x)e_i+\sum_{j\in J}\bar e_j(x)\bar e_j,\quad x\in X, \tag{22}$$

where the series converges unconditionally in H. Now observe that by (17) and (19), the ONS $(\bar e_j)_{j\in J}$ is actually an ONB of $\ker S_k^*$, and hence we have $[\bar e_j]_\sim=0$ for all $j\in J$. With a little extra effort, this leads to the following two versions of almost everywhere convergent Mercer representations.

Corollary 3.2 (Weak Mercer representation) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as in Lemma 2.12. Then for all $x \in X$, there exists an $N_x \subset X$ with $\nu(N_x) = 0$ such that*

$$k(x, x') = \sum_{i\in I} \mu_i e_i(x) e_i(x'), \quad x' \in X \setminus N_x, \tag{23}$$

where the series converges absolutely. In particular, the series converges $\nu \otimes \nu$-almost everywhere absolutely. Finally, if H is separable, then for ν-almost all $x \in X$, we have

$$k(\cdot, x) = \sum_{i\in I} \mu_i e_i(x) e_i,$$

where the series converges unconditionally in H.

Since convergence in H implies pointwise convergence, we see that, for all $x \in X$, the series in (22) evaluated at x converge to $k(x,x)$. Moreover, the occurring sums have nonnegative summands, and hence we conclude that the first series on the right-hand side of (22) converge on the diagonal $x = x'$. By Lemma 2.5, we thus find that the first series on the right-hand side of (22) actually converge not only on the diagonal but for all $x, x' \in X$. While, in general, the corresponding limit does not equal $k(x,x')$, it defines, however, a new kernel. The following result presents some properties of this kernel and its RKHS:

Theorem 3.3 (RKHS induced by a measure) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as in Lemma 2.12. Then the subset*

$$H_\nu := \overline{\operatorname{ran} S_k} = \overline{\operatorname{span}\{\sqrt{\mu_i} e_i : i \in I\}} = \overline{\operatorname{span}\{e_i : i \in I\}}$$

of H, equipped with the norm $\|f\|_{H_\nu} := \|f\|_H$ for all $f \in H_\nu$, is a separable RKHS over X. Moreover, $(\sqrt{\mu_i} e_i)_{i\in I}$ is an ONB of H_ν, and the kernel k_ν is given by

$$k_\nu(x, x') = \sum_{i\in I} \mu_i e_i(x) e_i(x'), \tag{24}$$

where the convergence is absolute for all $x, x' \in X$. In addition, we have $H_\nu^\perp = \{f \in H : [f]_\sim = 0\}$, $[H]_\sim = [H_\nu]_\sim$, and

$$[k(\cdot, x)]_\sim = [k_\nu(\cdot, x)]_\sim, \quad x \in X. \tag{25}$$

Finally, we have $H_\nu = H$, or equivalently $k_\nu = k$, if and only if $S_k^* : H \to L_2(\nu)$ is injective.

Note that the definitions $H_\nu := \overline{\operatorname{ran} S_k}$ and $\|f\|_{H_\nu} := \|f\|_H$ are independent of the chosen family $(e_i)_{i\in I} \subset H$. Consequently, the formula (24) is independent of the choice of $(e_i)_{i\in I} \subset H$, since every RKHS has a unique kernel. In other words, both H_ν and k_ν are uniquely determined for given H and ν. This justifies the notation H_ν and k_ν.

Theorem 3.3 shows, in particular, that the spaces H and H_ν are equal when interpreted as subsets of $L_2(\nu)$; that is, $[H]_\sim = [H_\nu]_\sim$. On the other hand, the kernel k_ν of the RKHS H_ν *always* has a pointwise convergent Mercer representation, while, in general, the original kernel k does not enjoy such a representation. Consequently, if we are *only* interested in the set $[H]_\sim$, it is often beneficial to consider the space H_ν instead, since for the latter we do have a Mercer representation. In this respect, the following corollary that computes the operators S_{k_ν}, $S^*_{k_\nu}$, and T_{k_ν} is often helpful, as we will see later on:

Corollary 3.4 *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Then for all $f \in L_2(\nu)$ and all $h \in H_\nu$, we have*

$$S_{k_\nu} f = S_k f, \tag{26}$$

$$S^*_{k_\nu} h = S^*_k h, \tag{27}$$

$$T_{k_\nu} f = T_k f. \tag{28}$$

Now observe that $S^*_{k_\nu}$ is injective by (27), the definition of H_ν, and (19). Consequently, we can equip the space $[H]_\sim = [H_\nu]_\sim = \operatorname{ran} S^*_{k_\nu}$ with the norm

$$\|f\|_{[H]_\sim} := \left\| \left(S^*_{k_\nu}\right)^{-1} f \right\|_H, \quad f \in [H]_\sim. \tag{29}$$

In other words, $[H]_\sim$ is the *space* that results from considering the RKHS H, or equivalently H_ν, as a subset of $L_2(\nu)$. Note that our construction in particular ensures that $[H]_\sim = [H_\nu]_\sim$ as spaces; that is, not only the sets are equal, but also their norms are equal.

The discussion above shows that we can consider the space $[H_\nu]_\sim$ whenever we are interested in properties of the space $[H]_\sim$. This observation will be used, e.g., in Theorem 4.6 and Corollary 5.5, where we investigate the approximation properties and the growth behavior of *infinite-sample* SVM solutions, respectively. Unfortunately, however, we will see below that a similar approach is, in general, *not* possible for the kernels k_ν and k. A typical example where we are interested in these kernels are *finite-sample* questions related to kernel-based statistical inference methods such as kernel PCA or SVMs. Indeed, these methods typically assume that there is a probability measure ν on X that generates observations $x_1, \dots, x_n \in X$ in an i.i.d. fashion, and the methods then use the information given by these observations with the help of the so-called Gram matrix $(k(x_i, x_j))_{i,j=1}^n$. Now, to replace this matrix by the Gram matrix of k_ν, we need to know that both matrices coincide, at least ν^n-almost surely. Clearly, the latter is satisfied, if k and k_ν coincide on a set of the form $(X \setminus N) \times (X \setminus N)$, where $\nu(N) = 0$, and we have already seen in Corollary 3.2 that such a set N exists for *separable* H. Our next goal is to provide another sufficient condition for the equality $k = k_\nu$ on a rectangular set of full measure.

Let us begin by recalling, see, e.g., [5, Chap. 25], that a Borel measure on a Hausdorff (topological) space (X,τ) is a measure ν on the Borel σ-algebra such that $\nu(A)<\infty$ for all compact $A\subset X$. Moreover, one can define the support of a Borel measure ν on (X,τ) by

$$\operatorname{supp}\nu := \bigl\{x\in X \mid \forall O\in\tau : x\in O \Rightarrow \nu(O)>0\bigr\}.$$

It is easy to see that $\operatorname{supp}\nu$ is closed and hence measurable. Moreover, for every open $N\subset X$ with $\nu(N)=0$, we have $N\subset X\setminus\operatorname{supp}\nu$. In particular, ν is strictly positive; that is, $\nu(O)>0$ for every nonempty open $O\subset X$ if and only if $\operatorname{supp}\nu=X$. For strictly positive Borel measures, we thus have $\nu(X\backslash\operatorname{supp}\nu)=0$. Moreover, if X has a countable base, we also have $\nu(X\backslash\operatorname{supp}\nu)=0$ by a simple argument that uses the fact that $X\setminus\operatorname{supp}\nu$ is open. Note that the latter observation in particular applies to Polish spaces. Analogously, if X is locally compact, that is, every $x\in X$ has a compact neighborhood, and ν is a Radon measure, i.e., an inner regular Borel measure, then $\nu(X\setminus\operatorname{supp}\nu)=0$, see [5, Chap. 28].

With these preparations, we can now investigate Mercer representations for continuous kernels.

Corollary 3.5 (Mercer representation for continuous kernels) *Let X be a Hausdorff space and ν be a Borel measure on X. Moreover, let $k: X\times X\to\mathbb{R}$ be a kernel whose RKHS H is compactly embedded into $L_2(\nu)$, and let $(e_i)_{i\in I}\subset H$ and $(\mu_i)_{i\in I}\subset(0,\infty)$ be as in Lemma 2.12. In addition, assume that one of the following two conditions is satisfied*:

(i) *k is continuous.*
(ii) *k is bounded and separately continuous*; *that is $k(\cdot,x)$ is continuous for all $x\in X$.*

Then for all fixed $x\in\operatorname{supp}\nu$, we have

$$k\bigl(x,x'\bigr)=\sum_{i\in I}\mu_i e_i(x)e_i\bigl(x'\bigr), \tag{30}$$

where the convergence is absolute for all $x'\in X$ and also uniform in x' on every subset $A\subset X$ for which we have $\sup_{x''\in A}k(x'',x'')<\infty$. In addition, if k is continuous, then the convergence in (30) is uniform in x and x' on all $A\times A\subset\operatorname{supp}\nu\times\operatorname{supp}\nu$ for which A is compact.

The corollary above shows that (30) converges absolutely for all $x,x'\in\operatorname{supp}\nu$. Consequently, if we are in a situation where $\nu(X\setminus\operatorname{supp}\nu)=0$, see the discussion in front of Corollary 3.5, then (30) holds on a set of the form $(X\setminus N)\times(X\setminus N)$, where $\nu(N)=0$. Moreover, if ν is a strictly positive measure, then (30) actually converges for *all* $x,x'\in X$. This leads to the following observation, which is interesting because, unlike in, e.g., [17, Lemmas 4.9 & 4.10], we do *not* assume that the underlying space X is separable.

Corollary 3.6 (Separability of RKHSs of continuous kernels) *Let X be a topological Hausdorff space and ν be a strictly positive Borel measure on X. Then every RKHS*

H that is compactly embedded into $L_2(\nu)$ and whose kernel k satisfies condition (i) *or* (ii) *of Corollary 3.5 is separable. In particular, if there exists a finite and strictly positive Borel measure ν on X, then every bounded and separately continuous kernel k has a separable RKHS H.*

Let us now return to the almost sure equality of the Gram matrices of k and k_ν. The following theorem shows that this equality can be guaranteed *if and only if* k and k_ν coincide on a rectangular set of full measure. It further shows that the latter is satisfied if and only if k and k_ν coincide almost everywhere on the diagonal.

Theorem 3.7 (Equality of Gram matrices) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Then the following statements are equivalent*:

(i) *There exists a measurable $N \subset X$ with $\nu(N) = 0$ such that $k(x, x') = k_\nu(x, x')$ for all $x, x' \in X \setminus N$.*
(ii) *For all $n \geq 1$, the Gram matrices $(k(x_i, x_j))_{i,j=1}^n$ and $(k_\nu(x_i, x_j))_{i,j=1}^n$ of k and k_ν coincide for ν^n-almost all $(x_1, \dots, x_n) \in X^n$.*
(iii) *There exists a measurable $N \subset X$ with $\nu(N) = 0$ such that $k(x, x) = k_\nu(x, x)$ for all $x \in X \setminus N$.*

One may be tempted to think that modifying the functions e_i on a set of measure zero would provide more cases in which the corresponding Gram matrices coincide. The following corollary shows that this is not the case:

Corollary 3.8 (Modifications of representing ONS) *Let X be a measurable space, ν be a probability measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i \in I} \subset H$ and $(\mu_i)_{i \in I} \subset (0, \infty)$ be as in Lemma 2.12. Then the following statements are equivalent*:

(i) *There exist a measurable $N \subset X$ with $\nu(N) = 0$, a family $(\bar{e}_i)_{i \in I}$ of functions with $[\bar{e}_i]_\sim = [e_i]_\sim$ for all $i \in I$, and a function $\bar{k} : X \times X \to \mathbb{R}$ satisfying*

$$\bar{k}(x, x') = \sum_{i \in I} \mu_i \bar{e}_i(x) \bar{e}_i(x'), \quad x, x' \in X \setminus N,$$

such that, for all $n \geq 1$, the Gram matrices $(k(x_i, x_j))_{i,j=1}^n$ and $(\bar{k}(x_i, x_j))_{i,j=1}^n$ coincide for ν^n-almost all $(x_1, \dots, x_n) \in X^n$.
(ii) *There exists a measurable $N \subset X$ with $\nu(N) = 0$ such that $k(x, x) = k_\nu(x, x)$ for all $x \in X \setminus N$.*

Summarizing the previous results, we see that, for Gram matrices, Mercer representations can be used if and only if k and k_ν coincide almost everywhere on the diagonal. Moreover, if the latter is satisfied, we can simply consider the Gram matrices of k_ν and the functions e_i in the Mercer representation. In addition, for separable RKHSs or (separately) continuous kernels on benign topological measure spaces, see the discussion around Corollary 3.5, such a replacement is always possible, while the next example shows that there do exist situations in which it is impossible.

Example 3.9 (No convergence on the diagonal for a nonseparable RKHS) Let $X = [0,1]$ and ν be the Lebesgue measure on X. Furthermore, we consider the kernel

$$k(x,x') := \mathbf{1}_{\{0\}}(x-x') = \sum_{x''\in[0,1]} \mathbf{1}_{\{x''\}}(x)\mathbf{1}_{\{x''\}}(x'), \quad x,x' \in X,$$

which equals 1 on the diagonal $x = x'$ and zero everywhere else. Consequently, we have $\|k\|_{\mathcal{L}_2(\nu)} = 1$, and hence the RKHS H of k is compactly embedded into $L_2(\nu)$. In addition, Remark 2.8 shows that H is given by

$$H := \left\{ \sum_{x\in X} a_x \mathbf{1}_{\{x\}} : (a_x) \in \ell_2(X) \right\},$$

and it is easy to check that the family $(\mathbf{1}_{\{x\}})_{x\in X}$ is an ONB of H. Moreover, the integral operator with respect to ν satisfies $T_k f = [0]_\sim$ for all $f \in L_2(\nu)$, and hence we have $I = \emptyset$ for the index set I in Lemma 2.12. By Theorem 3.3, we consequently obtain $k_\nu(x,x') = 0$ for all $x,x' \in X$, and hence k and k_ν do not coincide on a set of the form $(X\setminus N) \times (X\setminus N)$, where $\nu(N) = 0$. Furthermore, we actually have $k_\nu(x,x) \neq k(x,x)$ for *all* $x \in X$.

The example above presents an extreme case, in which the spaces H and $L_2(\nu)$ do not share any function except the zero function. In particular, it shows that the $\nu \otimes \nu$-almost everywhere convergence of Mercer representations of Corollary 3.2 cannot be improved in general. Moreover, for the pair H and $L_2(\nu)$ of the example, we still have $[H]_\sim = [H_\nu]_\sim$ as spaces, while the Gram matrices of k and k_ν *never* coincide. Consequently, without extra assumptions, the Gram matrices of k_ν are not suitable surrogates for the Gram matrices of k.

While the example above may seem to be somewhat artificial, it actually has some practical consequences for the analysis of learning algorithms. To see this, consider the so-called squared-hinge SVM, see [12, Chap. 6.1.2], which can be interpreted as a hard-margin SVM using the kernel $\tilde{k} := ck + k'$, where $c > 0$ is a constant derived from the regularization parameter of the squared-hinge SVM, k is the kernel of Example 3.9, and k' is the kernel used in the squared-hinge SVM. Now, our discussion above shows that in general we may not use $\tilde{k}_\nu$ in considerations on the Gram matrix of $\tilde{k}$, even if k' enjoys a pointwise convergent Mercer representation.

Our last goal of this section is to present an integral representation of self-adjoint nuclear operators $T : L_2(\nu) \to L_2(\nu)$. To this end, recall that nuclear operators are Hilbert-Schmidt, and hence there always exists a $k \in L_2(\nu \otimes \nu)$ with $T = T_k$, see, e.g., [45, Satz VI.6.3]. The following result establishes additional properties of this representing function k if T_k is positive. The general case, in which T is not necessarily positive, is discussed afterwards.

Theorem 3.10 (Mercer's theorem for positive self-adjoint nuclear operators) *Let X be a measurable space, ν be a measure on X, and $T : L_2(\nu) \to L_2(\nu)$ be a positive, self-adjoint, and nuclear operator. We write $(\mu_i)_{i\in I}$ for the at most countably many nonzero eigenvalues of T, where we included geometric multiplicities. Then there* *exists a measurable kernel k on X with separable RKHS H and $\|k\|_{\mathcal{L}_2(\nu)} < \infty$ such that*

$$Tf = \int_X k(\cdot, x') f(x') \, d\nu(x'), \quad f \in L_2(\nu), \tag{31}$$

where the left-hand side of the equation is considered to be a ν-equivalence class in $L_2(\nu)$. In addition, k enjoys a Mercer representation; that is, there exists an ONB $(\sqrt{\mu_i} e_i)_{i \in I}$ of H such that $([e_i]_\sim)_{i \in I}$ is an ONS in $L_2(\nu)$ consisting of the eigenfunctions corresponding to the eigenvalues $(\mu_i)_{i \in I}$ of T such that

$$k(x, x') = \sum_{i \in I} \mu_i e_i(x) e_i(x'), \quad x, x' \in X. \tag{32}$$

Moreover, the integral operator T_k implicitly used in (31) is defined pointwise; that is,

$$T_k f(x) := \int_X k(x, x') f(x') \, d\nu(x')$$

exists for all $f \in L_2(\nu)$ and all $x \in X$. Finally, the nuclear norm of T satisfies $\|T\|_{\mathsf{nuc}} = \|k\|^2_{\mathcal{L}_2(\nu)}$.

So far, Theorem 3.10 only provides additional properties of k if T is positive, self-adjoint, and nuclear. By the spectral theorem, however, general self-adjoint and nuclear operators $T : L_2(\nu) \to L_2(\nu)$ can be decomposed into two positive, self-adjoint, and nuclear operators T^+ and T^- with $T = T^+ - T^-$. Applying Theorem 3.10 to T^+ and T^-, we then see that $T = T_k$ for a function $k = k^+ - k^-$, where k^+ and k^- are kernels satisfying the properties listed in Theorem 3.10. Moreover, we have $\|T\|_{\mathsf{nuc}} = \|k^+\|^2_{L_2(\nu)} + \|k^-\|^2_{L_2(\nu)}$. Conversely, if we start with a difference of kernels $k := k^+ - k^-$, where $\|k^+\|^2_{L_2(\nu)} + \|k^-\|^2_{L_2(\nu)} < \infty$, then the associated integral operator T_k is self-adjoint and nuclear. In other words, self-adjoint, nuclear operators are exactly the operators that have an integral representation in terms of such a $k = k^+ - k^-$. Finally, if we start with a self-adjoint and nuclear integral operator $T_{\tilde{k}} : L_2(\nu) \to L_2(\nu)$, then the representing function k satisfies $k = \tilde{k}$ modulo a $\nu \otimes \nu$-zero set. Again, Example 3.9 can be used to show that, in general, we cannot expect that this zero set is the complement of a rectangular set.

4 Powers of Kernels and Integral Operators

In this section, we present properties of the fractional powers T_k^β of the integral operator T_k. We begin by investigating conditions that ensure that a fractional power is again an integral operator of a *kernel*. To this end, we introduce powers of kernels and RKHSs and provide sufficient conditions for their existence. In the second part of this section, we then present a formula for the image of the fractional power $T_k^{\beta/2}$ in terms of interpolation spaces.

Let us begin by recalling the (fractional) powers of the integral operators T_k. To this end, we consider the situation of Lemma 2.12. Then, for $\beta \in [0, \infty)$, the fractional power $T_k^\beta : L_2(\nu) \to L_2(\nu)$ is defined by

$$T_k^\beta f := \sum_{i\in I} \mu_i^\beta \langle f, [e_i]_\sim \rangle_{L_2(\nu)} [e_i]_\sim, \quad f \in L_2(\nu), \tag{33}$$

where we recall that this definition is actually independent of the chosen ONS of eigenvectors. Our first goal is to identify the fractional power T_k^β as an integral operator of a new kernel. To this end, we need to introduce powers of kernels and RKHSs.

Definition 4.1 (Powers of RKHSs) Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0, \infty)$ be as in Lemma 2.12. In addition, let $\beta > 0$ be such that

$$\sum_{i\in I} \mu_i^\beta e_i^2(x) < \infty, \quad x \in X. \tag{34}$$

Then we call the RKHS H_ν^β that is induced by (e_i) and (μ_i^β) the β-th power of H_ν. Moreover, its kernel k_ν^β is called the β-th power of k_ν.

Proposition 4.2 (Computing powers of RKHSs and kernels) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, for some fixed $\beta > 0$, let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0, \infty)$ be as in Definition 4.1. Then the RKHS H_ν^β is given by*

$$H_\nu^\beta = \left\{ \sum_{i\in I} a_i \mu_i^{\beta/2} e_i : (a_i) \in \ell_2(I) \right\} = \left\{ \sum_{i\in I} b_i e_i : (b_i) \in \ell_2\bigl(\mu^{-\beta}\bigr) \right\}$$

with the norm

$$\left\| \sum_{i\in I} b_i e_i \right\|_{H_\nu^\beta} := \bigl\| (b_i) \bigr\|_{\ell_2(\mu^{-\beta})}.$$

Furthermore, the RKHS H_ν^β is actually independent of the particular choice of $(e_i)_{i\in I} \subset H$ in Definition 4.1. Moreover, H_ν^β is separable and compactly embedded into $L_2(\nu)$. Finally, its kernel k_ν^β is given by

$$k_\nu^\beta\bigl(x, x'\bigr) = \sum_{i\in I} \mu_i^\beta e_i(x) e_i\bigl(x'\bigr), \quad x, x' \in X, \tag{35}$$

and again, this formula is independent of the chosen family $(e_i)_{i\in I} \subset H$ in Definition 4.1.

Since H_ν^β is compactly embedded into $L_2(\nu)$, we can consider H_ν^β as a subspace of $L_2(\nu)$. By the continuity of the embedding of H_ν^β into $L_2(\nu)$, we then obtain

$$\big[H_\nu^\beta\big]_\sim = \Big\{\sum_{i\in I} a_i \mu_i^{\beta/2}[e_i]_\sim : (a_i) \in \ell_2(I)\Big\} = \Big\{\sum_{i\in I} b_i[e_i]_\sim : (b_i) \in \ell_2\big(\mu^{-\beta}\big)\Big\}.$$

Now observe that even if (34) does not hold, the latter two sets can actually be defined because $(\mu_i)_{i\in I}$ is a bounded family and $([e_i]_\sim)_{i\in I}$ is an ONS of $L_2(\nu)$. In other words, if H is an RKHS that is compactly embedded into $L_2(\nu)$ and $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ are as in Lemma 2.12, then, for $\beta \geq 0$, we can always define

$$[H]_\sim^\beta := \Big\{\sum_{i\in I} a_i \mu_i^{\beta/2}[e_i]_\sim : (a_i) \in \ell_2(I)\Big\} = \Big\{\sum_{i\in I} b_i[e_i]_\sim : (b_i) \in \ell_2\big(\mu^{-\beta}\big)\Big\} \tag{36}$$

and equip this space with the Hilbert space norm

$$\Big\| \sum_{i\in I} b_i[e_i]_\sim \Big\|_{[H]_\sim^\beta} := \big\|(b_i)\big\|_{\ell_2(\mu^{-\beta})}.$$

Here we note that this norm is well defined, since $[H]_\sim^\beta \subset L_2(\nu)$ implies that every $h \in [H]_\sim^\beta$ has a *unique* representation in terms of the $L_2(\nu)$-ONS $([e_i]_\sim)_{i\in I}$. Moreover, it is easy to verify that $(\mu_i^{\beta/2}[e_i]_\sim)_{i\in I}$ is an ONB of $[H]_\sim^\beta$, and we will see in Theorem 4.6 that $[H]_\sim^\beta$ is, like H_ν^β, independent of the choice of $(e_i)_{i\in I} \subset H$ in Lemma 2.12.

The space $[H]_\sim^\beta$, which always exists, will be used later in this section to describe the images of the fractional powers of T_k, while powers of kernels and RKHSs, which do not necessarily exist, will be used in Sect. 5, where we investigate additional properties of these images.

Let us now return to our initial goal of this section, which was to find a representing kernel for the power T_k^β. The following theorem presents, besides some useful properties of the spaces H_ν^β, the first result in this direction:

Lemma 4.3 (Operators of powers of kernels) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, for some fixed $\beta > 0$, let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as Definition 4.1. Then we have*

$$T_k^\beta = T_{k_\nu^\beta}$$

and $[H_\nu^\beta]_\sim = [H]_\sim^\beta$. Moreover, for all $\alpha \geq \beta$, we have

$$\sum_{i\in I} \mu_i^\alpha e_i^2(x) < \infty, \qquad x \in X,$$

and the corresponding powers satisfy $H_\nu^\alpha \hookrightarrow H_\nu^\beta$.

Unfortunately, ensuring the summability (34) for *all* $x \in X$, is, in general, difficult. However, ensuring it only for ν-almost all $x \in X$ is often relatively easy, as we will see below. Consequently, let us now investigate how the construction of powers of kernels and RKHS can be suitably modified under this weaker assumption. To this end, assume that there exists a measurable $N \subset X$ with $\nu(N) = 0$ such that

$$\sum_{i \in I} \mu_i^\beta e_i^2(x) < \infty, \quad x \in X \backslash N. \tag{37}$$

We define $\bar{e}_i := \mathbf{1}_{X\backslash N} e_i$, $i \in I$, where $\mathbf{1}_A$ is the indicator function of a set A. Then we have $[\bar{e}_i]_\sim = [e_i]_\sim$ for all $i \in I$, and hence $([\bar{e}_i]_\sim)$ is an ONS in $L_2(\nu)$. Moreover, our construction yields

$$\sum_{i \in I} \mu_i^\beta \bar{e}_i^2(x) < \infty, \quad x \in X,$$

and hence $(\bar{e}_i)_{i \in I}$ and $(\mu_i^\beta)_{i \in I}$ induce an RKHS and a kernel, which we will denote by $\bar{H}_\nu^\beta$ and $\bar{k}_\nu^\beta$, respectively. From the construction of $\bar{H}_\nu^\beta$ in Lemma 2.6 and the fact that convergence in RKHSs implies pointwise convergence, it is easy to see that every $f \in \bar{H}_\nu^\beta$ satisfies $f(x) = 0$ for all $x \in N$. In addition, the construction guarantees

$$T_k^\beta = T_{\bar{k}_\nu^\beta}$$

and $[\bar{H}_\nu^\beta]_\sim = [H]_\sim^\beta$. In other words, the space $\bar{H}_\nu^\beta$ and its kernel $\bar{k}_\nu^\beta$ enjoy all the properties of the power H_ν^β and its kernel k_ν^β we have listed above. In general, however, the choice of N in (37) is not unique, and hence $\bar{H}_\nu^\beta$ is, unlike $[\bar{H}_\nu^\beta]_\sim$ and H_ν^β, not uniquely determined. Fortunately, it turns out that this ambiguity will not impact our results that involve $\bar{H}_\nu^\beta$, since we usually mention N explicitly. Finally, if (37) actually holds for *all* $x \in X$, we obtain $[H_\nu^\beta]_\sim = [\bar{H}_\nu^\beta]_\sim$.

Above, we have motivated the construction of $\bar{H}_\nu^\beta$ by the claim that, in general, it is easier to guarantee (37) than (34). The following two results support this claim by providing sufficient conditions for (37):

Proposition 4.4 (Eigenvalue summability ensures existence of powers) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i \in I} \subset H$ and $(\mu_i)_{i \in I} \subset (0, \infty)$ be as in Lemma 2.12. Then for $\beta > 0$, the following statements are equivalent*:

(i) *The family $(\mu_i)_{i \in I}$ is β-summable; that is, $\sum_{i \in I} \mu_i^\beta < \infty$.*
(ii) *There exists a measurable $N \subset X$ with $\nu(N) = 0$ such that (37) holds and the corresponding induced kernel $\bar{k}_\nu^\beta$ satisfies $\|\bar{k}_\nu^\beta\|_{\mathcal{L}_2(\nu)} < \infty$.*

If one (and thus both) statement is true, we further have $\|\bar{k}_\nu^\beta\|_{\mathcal{L}_2(\nu)}^2 = \sum_{i \in I} \mu_i^\beta$.

Sometimes even the β-summability of the eigenvalues is hard to check. In this case, the following result provides another way to ensure that (37) holds ν-almost surely. This result is particularly useful in combination with a later result of this section that describes $[H]_\sim^\beta$ by an interpolation space.

Proposition 4.5 (Boundedness ensures existence of powers) *Let* $(X, \mathcal{A})$ *be a measurable space and* ν *be a* σ*-finite measure on* $(X, \mathcal{A})$ *such that* $\mathcal{A}$ *is* ν*-complete. Moreover, let k be a measurable kernel on X whose RKHS H is compactly embedded into* $L_2(\nu)$. *Assume that, for some* $\beta > 0$, *we have* $[H]_\sim^\beta \hookrightarrow L_\infty(\nu)$. *Then there exists a measurable* $N \subset X$ *with* $\nu(N) = 0$ *such that* (37) *holds.*

For finite measures, we will see in Theorem 5.3, that the continuous inclusion $[H]_\sim^\beta \hookrightarrow L_\infty(\nu)$ implies the β-summability of the eigenvalues $(\mu_i)_{i\in I}$. In other words, for such ν, the condition in Proposition 4.5 is actually stronger than condition (i) in Proposition 4.4. On the other hand, we will see below that $[H]_\sim^\beta$ equals a certain interpolation space, and sometimes it may be easier to investigate the properties of this interpolation space than the behavior of the eigenvalues, when one wants to ensure the existence of $\bar{k}_\nu^\beta$.

Let us now turn to the second goal of this section, namely, the description of $[H]_\sim^\beta$ and the range of T_k^β in terms of interpolation spaces between H and $L_2(\nu)$. Before we can do this, we need to briefly recall the definition of interpolation spaces of the real method. To this end, we fix two arbitrary Banach spaces E_0 and E_1 that are continuously embedded in some topological (Hausdorff) vector space $\mathcal{E}$. Then, for $x \in E_0 + E_1$ and $t > 0$, the K-functional of the real interpolation method is defined by

$$\begin{aligned} K(x,t) &:= K(x,t,E_0,E_1) \\ &:= \inf\bigl\{\|x_0\|_{E_0} + t\|x_1\|_{E_1} : x_0 \in E_0, x_1 \in E_1, x = x_0 + x_1\bigr\}. \end{aligned}$$

It is not hard to see that in the case of two Banach spaces E and F satisfying $F \hookrightarrow E$, the K-functional of $x \in E$ is given by

$$K(x,t,E,F) = \inf_{y\in F}\bigl(\|x-y\|_E + t\|y\|_F\bigr), \qquad t > 0.$$

With the help of the K-functional, one can define interpolation norms, for $0 < \beta < 1$, $1 \le r \le \infty$, and $x \in E_0 + E_1$, by

$$\|x\|_{\beta,r} := \begin{cases} (\int_0^\infty (t^{-\beta}K(x,t))^r t^{-1}\,dt)^{1/r} & \text{if } 1 \le r < \infty, \\ \sup_{t>0} t^{-\beta}K(x,t) & \text{if } r = \infty. \end{cases}$$

Moreover, the corresponding interpolation spaces, see [6, Definition 1.7 on p. 299], are defined by

$$[E_0, E_1]_{\beta,r} := \bigl\{x \in E_0 + E_1 : \|x\|_{\beta,r} < \infty\bigr\}.$$

With the help of these preparations, we can now present the main result of this section that describes the image of the fractional power $T_k^{\beta/2}$ by a suitable interpolation space.

Theorem 4.6 (Images of powers are interpolation spaces) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Then for all $0 < \beta < 1$, we have*

$$\operatorname{ran} T_k^{\beta/2} = [H]_\sim^\beta = \big[L_2(\nu), [H]_\sim\big]_{\beta,2},$$

and the spaces $[H]_\sim^\beta$ and $[L_2(\nu), [H]_\sim]_{\beta,2}$ have equivalent norms. Moreover, there exist constants $c_\beta, C_\beta \in (0, \infty)$ such that

$$c_\beta \|f\|_{L_2(\nu)} \le \big\|T_k^{\beta/2} f\big\|_{[L_2(\nu),[H]_\sim]_{\beta,2}} \le C_\beta \|f\|_{L_2(\nu)}$$

for all $f \in \overline{\operatorname{span}\{[e_i]_\sim : i \in I\}}$, where $(e_i)_{i\in I} \subset H$ is as in Lemma 2.12.

It has been shown in [34] that the interpolation spaces $[L_2(\nu), [H]_\sim]_{\beta,\infty}$ exactly describe the approximation properties of LS-SVMs. On the other hand, several papers dealing with learning rates for these SVMs, see, e.g., [9, 15], describe their approximation properties in terms of $\operatorname{ran} T_k^\beta$. So far, the only link between these two descriptions is an "almost equivalence", see, e.g., [14, Theorem 4.1], which holds for compact metric spaces X; continuous kernels k; and strictly positive, finite measures ν. Without making these quite restrictive assumptions, Theorem 4.6 now identifies the image $\operatorname{ran} T_k^\beta$ with another interpolation space, namely, $[L_2(\nu), [H]_\sim]_{\beta,2}$. Since this interpolation space sits in between spaces of the scale $[L_2(\nu), [H]_\sim]_{p,\infty}$, $p \in (0, 1)$, we immediately obtain the following corollary, which generalizes the above mentioned [14, Theorem 4.1] to basically arbitrary input spaces X, kernels k, and measures ν.

Corollary 4.7 *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Then for all $0 < \beta < 1$ and all $\varepsilon > 0$ with $\beta + \varepsilon < 1$, we have*

$$\big[L_2(\nu), [H]_\sim\big]_{\beta+\varepsilon,\infty} \subset \operatorname{ran} T_k^{\beta/2} \subset \big[L_2(\nu), [H]_\sim\big]_{\beta,\infty}.$$

Theorem 4.6 and its corollary have an immediate influence on some recent articles, such as [8, 9, 38], that establish learning rates for LS-SVMs. Indeed, authors have typically restricted their considerations to target functions contained either in some $\operatorname{ran} T_k^{\beta/2}$ or in some $[L_2(\nu), [H]_\sim]_{\beta,\infty}$. So far, however, these assumption could not be compared for general X and k, which is now possible due to Corollary 4.7.

5 Bounded Powers of Kernels and an Application to SVMs

In this section, we investigate in which situations the space $[H]_\sim^\beta$, or equivalently, the space $[L_2(\nu), [H]_\sim]_{\beta,2}$, is continuously embedded into $L_\infty(\nu)$. In particular, we will show that such a continuous embedding implies the β-summability of the eigenvalues. In the second part of this section, we will then present an application that deals with an important aspect in the statistical analysis of support vector machines. Here, we will present a sharper bound on one of the key quantities of these learning methods, and we will also briefly discuss possible implications.

We begin with the following lemma that characterizes RKHSs that are continuously embedded into $L_\infty(\nu)$. While this result does not seem to be surprising at all, its proof is not trivial, since, unlike for the wellknown characterization of $H \hookrightarrow \ell_\infty(X)$, see, e.g., [36, Lemma 4.23], we have to carefully deal with zero sets.

Lemma 5.1 (Almost surely bounded kernels) *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel whose RKHS H is separable. Then the following statements are equivalent*:

(i) *The map $[\,\cdot\,]_\sim : H \to L_\infty(\nu)$ is well defined and continuous.*
(ii) *There exist a constant $\kappa \in [0,\infty)$ and a measurable $N \subset X$ with $\nu(N) = 0$ such that, for all $f \in H$, we have*

$$\bigl|f(x)\bigr| \le \kappa\,\|f\|_H, \quad x \in X\backslash N.$$

(iii) *There exist a constant $\kappa \in [0,\infty)$ and a measurable $N \subset X$ with $\nu(N) = 0$ and*

$$k(x,x) \le \kappa^2, \quad x \in X\backslash N. \tag{38}$$

Moreover, the same equivalence holds if X is a topological Hausdorff space, ν is a Borel measure on X satisfying $\nu(X\backslash \operatorname{supp}\nu) = 0$, and $k : X \times X \to \mathbb{R}$ is a kernel as in Corollary 3.5.

It is easy to check that the implications (iii) ⇔ (ii) ⇒ (i) actually hold for all RKHSs that have a measurable kernel. However, the remaining implication (i) ⇒ (ii) requires additional assumptions, since for the proof of (ii), or equivalently, (iii), we need to control *uncountable* unions of ν-zero sets. To raise caution in this regard, the following example, which is a modification of Example 3.9, shows that for nonseparable RKHSs, the implication (i) ⇒ (ii) is, in general, not true, even if the RKHS, is compactly embedded into $L_2(\nu)$.

Example 5.2 (Almost surely vanishing functions but no almost surely bounded kernel) Let $X = (0,1]$, and let ν be the Lebesgue measure on X. Furthermore, we consider the kernel

$$k\bigl(x,x'\bigr) := \sum_{x'' \in (0,1]} \frac{\mathbf{1}_{\{x''\}}(x)\mathbf{1}_{\{x''\}}(x')}{\sqrt{x''}}, \quad x,x' \in X,$$

which equals $x^{-1/2}$ on the diagonal $x = x'$ and zero everywhere else. Consequently, we have $\|k\|_{\mathcal{L}_2(\nu)} = \sqrt{2}$, and hence the RKHS H of k is compactly embedded into $L_2(\nu)$. In addition, H is given by

$$H := \biggl\{\sum_{x\in X} a_x x^{-1/4}\,\mathbf{1}_{\{x\}} : (a_x) \in \ell_2(X)\biggr\},$$

and the family $(x^{-1/4}\mathbf{1}_{\{x\}})_{x\in X}$ is an ONB of H. Moreover, every $(a_x)\in\ell_2(X)$ satisfies $a_x=0$ for all but at most countably many $x\in X$, and hence we have $[f]_\sim=0$ for all $f\in H$. Clearly, this shows the continuity of $[\,\cdot\,]_\sim: H\to L_\infty(\nu)$, but, of course, k is not ν-almost surely bounded on the diagonal.

With the help of these preparations, we can now establish the following result that characterizes powers of RKHSs that consist of bounded functions:

Theorem 5.3 (Bounded powers of kernels) *Let $(X,\mathcal{A})$ be a measurable space, ν be a σ-finite measure on $(X,\mathcal{A})$ such that $\mathcal{A}$ is ν-complete, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Furthermore, let $(e_i)_{i\in I}\subset H$ and $(\mu_i)_{i\in I}\subset(0,\infty)$ be as in Lemma 2.12. Then, for all $0<\beta\le 1$, the following statements are equivalent*:

(i) *$[H]_\sim^\beta$ is continuously embedded into $L_\infty(\nu)$; that is, $[H]_\sim^\beta\hookrightarrow L_\infty(\nu)$.*
(ii) *There exist a measurable $N\subset X$ with $\nu(N)=0$ and a constant $\kappa\in[0,\infty)$ such that*

$$\sum_{i\in I}\mu_i^\beta e_i^2(x)\le\kappa^2,\qquad x\in X\backslash N.$$

(iii) *There exists a constant $C\in[0,\infty)$ such that $\|T_k^{\beta/2}f\|_{L_\infty(\nu)}\le C\,\|f\|_{L_2(\nu)}$ for all $f\in L_2(\nu)$.*

Moreover, if one (and thus all) of the statements above is true, there exists a constant $c\in[0,\infty)$ such that

$$\|f\|_{L_\infty(\nu)}\le c\,\|f\|_{[H]_\sim}^{\beta}\,\|f\|_{L_2(\nu)}^{1-\beta},\qquad f\in[H]_\sim. \tag{39}$$

Finally, assume that ν is a finite measure. Then $[H]_\sim^\beta\hookrightarrow L_\infty(\nu)$ implies $\sum_{i\in I}\mu_i^\beta<\infty$, and the weaker Condition (39) implies $\sum_{i\in I}\mu_i^{\beta+\varepsilon}<\infty$ for all $\varepsilon>0$.

It has been recently discovered in [20, 37] that the statistical analysis of SVMs can be tightened for RKHSs that allow a refined control of the $\|\cdot\|_\infty$-norm of their functions in the sense of (39). So far, however, only little has been known about sufficient conditions for (39). Namely, [20] showed that, for *uniformly bounded* representatives $(e_i)_{i\in I}$, that is, $\sup_{i\in I}\|e_i\|_\infty<\infty$, the condition $\mu_i\le c\,i^{-\beta}$, $i\in I$, where $c\in[0,\infty)$ is a constant independent of i, implies (39). Unfortunately, however, it is known, see [47, Example 1], that even C^∞-kernels on $[0,1]$, where $[0,1]$ is equipped with the Lebesgue measure, may not have uniformly bounded representatives $(e_i)_{i\in I}$ of the eigenfunctions, and, in general, the uniform boundedness is hard to check. Moreover, even for bounded kernels, the condition $\mu_i\le ci^{-\beta}$, $i\in I$, is not sufficient for (ii) of Theorem 5.3. Indeed, [47, Example 2] shows that there exists a continuous kernel on $[0,1]$, where $[0,1]$ is equipped with the Lebesgue measure, such that $\mu_i\le c\,i^{-2}8^{-i}$ for a constant $c\in[0,\infty)$ and all $i\ge 1$, but (ii) of Theorem 5.3 fails to be true for $\beta=1/2$. Another sufficient condition for (39) was used in [37]. Namely, [37] used [6, Proposition 2.10 on page 316] that shows that (39) is satisfied, if and only if $[L_2(\nu),[H]_\sim]_{\beta,1}\hookrightarrow L_\infty(\nu)$. Moreover, they illustrated the latter situation, when H is a Sobolev space of sufficient smoothness. However, [37] was not aware of the fact that the slightly stronger condition $[L_2(\nu), [H]_\sim]_{\beta,2} \hookrightarrow L_\infty(\nu)$, which was implicitly used in [37] in form of (iii), actually implies the β-summability of the eigenvalues, whenever the measure is finite.

Our last goal is to illustrate the consequences of the developed theory for the analysis of SVMs. To this end, let $L : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ be the least-squares loss defined by $L(y,t) := (y-t)^2$, $y, t \in \mathbb{R}$. Moreover, let P be distribution on $X \times \mathbb{R}$ with finite average 2nd moment, that is,

$$|\mathrm{P}|_2^2 := \int_{X\times Y} |y|^2 \, d\mathrm{P}(x,y) < \infty.$$

In the following, we always consider the probability measure $\nu := \mathrm{P}_X$ on X; that is, ν equals the marginal distribution of P on X. For an $f \in \mathcal{L}_2(\nu)$, we define the least-squares risk by

$$\mathcal{R}_{L,\mathrm{P}}(f) := \int_{X\times Y} L\bigl(y - f(x)\bigr) \, d\mathrm{P}(x,y) = \int_{X\times Y} \bigl(y - f(x)\bigr)^2 \, d\mathrm{P}(x,y),$$

and, in addition, the Bayes least-squares risk is defined by $\mathcal{R}^*_{L,\mathrm{P}} := \inf\{\mathcal{R}_{L,\mathrm{P}}(f) : f \in \mathcal{L}_2(\nu)\}$. It is well known that this infimum is attained by the function $x \mapsto \mathbb{E}_\mathrm{P}(Y|x)$, $x \in X$, where we note that the regular conditional expectation $\mathbb{E}_\mathrm{P}(Y|x)$ is P_X-almost surely unique. In the following, we write $f^*_{L,\mathrm{P}} := [x \mapsto \mathbb{E}_\mathrm{P}(Y|x)]_\sim$ for the equivalence class of the conditional expectation. Note that actually every $f \in [x \mapsto \mathbb{E}_\mathrm{P}(Y|x)]_\sim$ minimizes the least-squares risk.

LS-SVMs try to approximate the minimizer $f^*_{L,\mathrm{P}}$ by minimizing a regularized (empirical) risk, or, more precisely, they solve the optimization problem

$$\arg\min_{f\in H} \bigl(\lambda \|f\|_H^2 + \mathcal{R}_{L,\mathrm{Q}}(f)\bigr),$$

where H is an RKHS of a bounded measurable kernel, $\lambda > 0$ is a regularization parameter, and Q is an (empirical) distribution obtained from some observations $((x_1,y_1), \ldots, (x_n,y_n)) \in (X \times \mathbb{R})^n$. It turns out that this optimization problem has a unique solution $f_{\mathrm{Q},\lambda} \in H$ for all distributions Q on $X \times \mathbb{R}$ with $|\mathrm{Q}|_2^2 < \infty$, see [36, Chap. 5.1].

Although LS-SVMs only solve the optimization problem for *empirical* distributions, the statistical analysis of LS-SVMs also uses the "infinite-sample solution" $f_{\mathrm{P},\lambda}$, see, e.g., [36, Chap. 7.4] and [37]. In particular, bounds for $\|f_{\mathrm{P},\lambda}\|_\infty$ are crucial in the analysis. In the following, we show how the developed theory can be used to improve such bounds. We begin with a result on the Fourier coefficients of $f_{\mathrm{P},\lambda}$ that generalizes an analogous formula from [46] for continuous kernels on compact metric spaces.

Theorem 5.4 (Fourier coefficients of SVMs) *Let X be a measurable space*, P *be a probability measure on $X \times [-1,1]$, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$, where $\nu := \mathrm{P}_X$. Furthermore, let $(e_i)_{i\in I} \subset H$*

and $(\mu_i)_{i\in I} \subset (0,\infty)$ *be as in Lemma* 2.12. *Then we have* $f_{\mathrm{P},\lambda} \in H_\nu$ *and*

$$\big\langle f_{\mathrm{P},\lambda}, \sqrt{\mu_i}e_i\big\rangle_H = \frac{\sqrt{\mu_i}}{\lambda+\mu_i}\big\langle f^*_{L,\mathrm{P}}, [e_i]_\sim\big\rangle_{L_2(\nu)}, \quad i \in I.$$

Note that there is a low pass filter interpretation for the formula provided in Theorem 5.4, see [46] and the references therein.

Combining Theorem 5.4 with Theorem 5.3, we can now establish the sharpest known supremum bound on $f_{\mathrm{P},\lambda}$.

Corollary 5.5 (Improved supremum bound on SVMs) *Let* X *be a measurable space*, P *be a probability measure on* $X\times[-1,1]$, *and* k *be a measurable kernel on* X *whose RKHS* H *is compactly embedded into* $L_2(\nu)$, *where* $\nu := \mathrm{P}_X$. *Assume that, for some* $\alpha \in (0,1]$, *we have* $[H]^\alpha_\sim \hookrightarrow L_\infty(\nu)$. *Moreover, assume that, for some* $\beta \in (0,\alpha]$, *we have* $f^*_{L,\mathrm{P}} \in [H]^\beta_\sim = \operatorname{ran} T_k^{\beta/2}$. *Then there exists a constant* $c \in [0,\infty)$ *such that*

$$\big\|[f_{\mathrm{P},\lambda}]_\sim\big\|_{L_\infty(\nu)} \le c\,\lambda^{-\frac{\alpha-\beta}{2}}, \quad \lambda \in (0,1].$$

Recall that, for bounded kernels and distributions P on $X\times[-1,1]$, we always obtain, see [36, Chap. 5.1], the bound

$$\big\|[f_{\mathrm{P},\lambda}]_\sim\big\|_{L_\infty(\nu)} \le c\,\lambda^{-\frac{1}{2}}, \quad \lambda \in (0,1].$$

Moreover, if $f^*_{L,\mathrm{P}} \in [L_2(\nu),[H]_\sim]_{\beta,\infty}$, this can be easily improved, see [36, Chapter 5.4], to

$$\big\|[f_{\mathrm{P},\lambda}]_\sim\big\|_{L_\infty(\nu)} \le c\,\lambda^{-\frac{1-\beta}{2}}, \quad \lambda \in (0,1].$$

This bound was further improved in [37], where, compared to Corollary 5.5, the slightly weaker assumptions $[L_2(\nu),[H]_\sim]_{\alpha,1} \hookrightarrow L_\infty(\nu)$ and $f^*_{L,\mathrm{P}} \in [L_2(\nu),[H]_\sim]_{\beta,\infty}$ were used to establish the bound

$$\big\|[f_{\mathrm{P},\lambda}]_\sim\big\|_{L_\infty(\nu)} \le c\,\lambda^{-\frac{\alpha-\alpha\beta}{2}}, \quad \lambda \in (0,1]. \tag{40}$$

On the other hand, the continuous embeddings $[L_2(\nu),[H]_\sim]_{\alpha+\varepsilon,2} \hookrightarrow [L_2(\nu),[H]_\sim]_{\alpha+\varepsilon,\infty} \hookrightarrow [L_2(\nu),[H]_\sim]_{\alpha,1}$ and $[L_2(\nu),[H]_\sim]_{\beta,\infty} \hookrightarrow [L_2(\nu),[H]_\sim]_{\beta-\varepsilon,1} \hookrightarrow [L_2(\nu),[H]_\sim]_{\beta-\varepsilon,2}$, which hold for all sufficiently small $\varepsilon > 0$, see Lemma 6.5, show that Corollary 5.5 yields the bound

$$\big\|[f_{\mathrm{P},\lambda}]_\sim\big\|_{L_\infty(\nu)} \le c_\varepsilon\,\lambda^{-\frac{\alpha-\beta}{2}-\varepsilon}, \quad \lambda \in (0,1], \tag{41}$$

whenever the conditions of [37] are satisfied. For the most interesting case $\alpha < 1$, the bound (41) is strictly sharper than that of [37], since we can choose an arbitrarily small $\varepsilon > 0$. On the other hand, (40) has been used in [37] to obtain the sharpest known (and often also asymptotically optimal) learning rates for LS-SVMs. With some extra effort and the improved bound (41), it is now possible to establish these learning rates under assumptions on H and P that are significantly milder than those of [37]. Since this is clearly out of the scope of this paper, we refer to a forthcoming paper.

6 Proofs

This section contains the proofs of the theorems presented in this work. The structure of this section follows the structure of the previous sections; that is, the proofs for each of the Sects. 2 to 5 are contained in a separate subsection.

6.1 Proofs Related to Compactly Embedded RKHSs

Proof of Lemma 2.2 The adjoint $I_k^* : L_2(\nu) \to H$ of $I_k : H \to L_2(\nu)$ is defined by the relation

$$\langle I_k^* f, h\rangle_H = \langle f, I_k h\rangle_{L_2(\nu)} = \int_X f h \, d\nu, \qquad f \in L_2(\nu), h \in H.$$

For $x \in X$ and $h := k(\,\cdot\,, x)$, this yields

$$I_k^* f(x) = \langle I_k^* f, k(\,\cdot\,, x)\rangle_H = \int_X k(x, x') f(x') \, d\nu(x'),$$

and hence we have shown (7). From this, the factorization easily follows, and thus T_k is continuous and positive. Moreover, the self-adjointness and compactness also immediately follow from this factorization. □

Proof of Lemma 2.3 For $f \in H$, we have

$$\begin{aligned}\int_X |f(x)|^2 \, d\nu(x) &= \int_X \left|\langle f, k(\,\cdot\,, x)\rangle\right|^2 d\nu(x) \le \|f\|_H^2 \int_X \left\|k(\,\cdot\,, x)\right\|_H^2 d\nu(x) \\ &= \|f\|_H^2 \, \|k\|_{\mathcal{L}_2(\nu)}^2,\end{aligned}$$

where in the last step we used $\|k(\,\cdot\,, x)\|_H^2 = k(x, x)$ for all $x \in X$. Consequently, H is embedded into $L_2(\nu)$. Now let $(e_j)_{j\in J}$ be a not necessarily countable ONB of H. For finite $A \subset J$, we then have

$$\begin{aligned}\sum_{j\in A} \|S_k^* e_j\|_{L_2(\nu)}^2 &= \sum_{j\in A} \int_X |S_k^* e_j|^2 \, d\nu = \int_X \sum_{j\in A} |e_j(x)|^2 \, d\nu(x) \\ &= \int_X \sum_{j\in A} \left|\langle e_j, k(\,\cdot\,, x)\rangle_H\right|^2 d\nu(x) \\ &\le \int_X \left\|k(\,\cdot\,, x)\right\|_H^2 d\nu(x) \\ &= \|k\|_{\mathcal{L}_2(\nu)}^2,\end{aligned}$$

where for the inequality we used Bessel's inequality. From this, it is easy to conclude that $S_k^* = I_k$ is Hilbert-Schmidt, and hence so is S_k. Consequently, S_k, I_k, and T_k are compact, and, as a product of two Hilbert-Schmidt operators, T_k is even nuclear.

The last assertion follows from $|k(x,x')|^2 \le k(x,x)k(x',x')$ for all $x, x' \in X$, see, e.g., [36, p. 124]. □

Proof of Lemma 2.4 Without loss of generality, we may assume that $\sup_{j\in J} \|[e_j]_\sim\|_{L_2(\nu)} \le 1$. Let us fix an $f \in H$. Then there exists an $(a_j) \in \ell_2(J)$ with $f = \sum_{j\in J} a_j e_j$, and there further exists an at most countable $I \subset J$ with $a_j = 0$ for all $j \in J \setminus I$. Consequently, we have $f = \sum_{i\in I} a_i e_i$, where the convergence is unconditional in H. If I is finite, we may assume without loss of generality that $I = \{1, \dots, n\}$, and a simple calculation shows

$$\begin{aligned}\int_X \bigl|f(x)\bigr|^2 \, d\nu(x) &= \int_X \left|\sum_{i=1}^n a_i e_i(x)\right|^2 d\nu(x) = \sum_{i,j=1}^n a_i a_j \int_X e_i(x) e_j(x) \, d\nu(x) \\ &= \sum_{i,j=1}^n a_i a_j \bigl\langle [e_i]_\sim, [e_j]_\sim \bigr\rangle_{L_2(\nu)} \\ &\le \sum_{i=1}^n a_i^2.\end{aligned}$$

By Parseval's identity, we then obtain $\|[f]_\sim\|_{L_2(\nu)} \le \|f\|_H$. Let us now assume that I is not finite. Since I is countable, we then may assume without loss of generality that $I = \mathbb{N}$. Now recall that convergence in H implies pointwise convergence, and thus Fatou's Lemma together with calculations analogous to those above implies

$$\begin{aligned}\int_X \bigl|f(x)\bigr|^2 \, d\nu(x) &= \int_X \liminf_{n\to\infty} \left|\sum_{i=1}^n a_i e_i(x)\right|^2 d\nu(x) \le \liminf_{n\to\infty} \int_X \left|\sum_{i=1}^n a_i e_i(x)\right|^2 d\nu(x) \\ &\le \liminf_{n\to\infty} \sum_{i=1}^n a_i^2 \le \|f\|_H^2.\end{aligned}$$

To show that I_k is compact, we fix a $\delta > 0$ and define

$$J_\delta := \bigl\{ j \in J : \bigl\|[e_j]_\sim\bigr\|_{L_2(\nu)} > \delta \bigr\}.$$

By our assumption, we then see that J_δ is a finite set. Let us further denote the orthogonal projection of H onto $\overline{\operatorname{span}\{e_j : j \in J_\delta\}}$ by P_δ. Obviously, P_δ has finite rank. Moreover, for $f \in H$, the continuity of I_k and the orthogonality of the $[e_j]_\sim$ yield

$$\begin{aligned}\|I_k P_\delta f - I_k f\|_{L_2(\nu)}^2 &= \left\| I_k\biggl(\sum_{j\notin J_\delta} \langle f, e_j\rangle_H e_j \biggr) \right\|_{L_2(\nu)}^2 = \left\| \sum_{j\notin J_\delta} \langle f, e_j\rangle_H [e_j]_\sim \right\|_{L_2(\nu)}^2 \\ &\le \sum_{j\notin J_\delta} \bigl|\langle f, e_j\rangle_H\bigr|^2 \delta^2 \le \delta^2 \|f\|_H^2.\end{aligned}$$

Consequently, I_k can be approximated by the finite rank operators $I_k P_\delta$, $\delta > 0$, and hence it is compact. □

For the proof of Lemma 2.5, we need the following auxiliary result:

Lemma 6.1 *Let $(b_i)_{i\geq 1}$ be a sequence of real numbers such that the series*

$$\sum_{i=1}^{\infty} a_i b_i \tag{42}$$

converges for all $(a_i)\in \ell_2$. Then we have $(b_i)\in \ell_2$.

Proof We first show that the series in (42) converges absolutely for all $(a_i)\in\ell_2$. To this end, we fix an $(a_i)\in\ell_2$ and define the sequence $(\bar a_i)$ by $\bar a_i := a_i$ if $a_i b_i \geq 0$ and $\bar a_i := -a_i$ if $a_i b_i < 0$. This clearly gives $\bar a_i b_i = |a_i b_i|$ for all $i\in I$ and $(\bar a_i)\in\ell_2$, and hence $\sum_{i=1}^\infty |a_i b_i| = \sum_{i=1}^\infty \bar a_i b_i$ converges. For $n\geq 1$, we now consider the bounded linear operators

$$\begin{aligned} S_n : \ell_2 &\to \mathbb{R},\\ (a_i) &\mapsto \sum_{i=1}^n a_i b_i. \end{aligned}$$

Then an elementary calculation using the equality condition in Hölder‘s inequality shows $\|S_n\| = (\sum_{i=1}^n b_i^2)^{1/2}$ for all $n\geq 1$. Moreover, for fixed $a := (a_i)\in\ell_2$, we have

$$\sup_{n\geq 1}\|S_n a\|_{\mathbb{R}} = \sup_{n\geq 1}\left|\sum_{i=1}^n a_i b_i\right| \leq \sup_{n\geq 1}\sum_{i=1}^n |a_i b_i| = \sum_{i=1}^\infty |a_i b_i| < \infty.$$

Applying the Theorem of Banach-Steinhaus, see, e.g., [11, Corollary 14.3, p. 96] or [45, Theorem IV.2.1], then yields $\sup_{n\geq 1}\|S_n\| < \infty$. By the formula for $\|S_n\|$ established above, we thus obtain the assertion. □

Proof of Lemma 2.5 If I is finite, there is nothing to prove, and hence we assume that $I = \mathbb{N}$. The equivalence of (i) and (ii) follows from Lemma 6.1 and Hölder's inequality. Moreover, by considering $x' = x$, we see that (iii) implies (ii), and the converse implication can be shown by Hölder's inequality. Finally, the absolute convergence in (i) and (iii) follows from another application of Hölder's inequality. □

For the proof of Lemma 2.6 we need the following result, which characterizes ONBs of RKHSs in terms of their ability to represent the kernel:

Theorem 6.2 *Let H be an RKHS with kernel $k : X\times X\to\mathbb{R}$, and $(e_j)_{j\in J}$ be an ONS in H. Then the following statements are equivalent*:

(i) *$(e_j)_{j\in J}$ is an ONB of H.*
(ii) *For all $x\in X$, we have the following unconditional convergence in H*:

$$k(\cdot,x) = \sum_{j\in J} e_j(x) e_j.$$

Proof (i) $\Rightarrow$ (ii). Using the ONB representation of $k(\cdot,x)\in H$ and the reproducing property, we obtain

$$k(\cdot,x)=\sum_{j\in J}\big\langle k(\cdot,x),e_j\big\rangle_H e_j=\sum_{j\in J}e_j(x)e_j,$$

where the series converges, as every Fourier series in a Hilbert space, unconditionally.

(ii) $\Rightarrow$ (i). Obviously, it suffices to show that $\operatorname{span}\{e_j : j\in J\}$ is dense in H. To this end, we first recall, see, e.g., [36, Theorem 4.21], that

$$H=\overline{\operatorname{span}\big\{k(\cdot,x):x\in X\big\}}.$$

Moreover, (ii) ensures that $k(\cdot,x)\in\overline{\operatorname{span}\{e_j : j\in J\}}$, from which we conclude

$$H\subset\overline{\operatorname{span}\overline{\operatorname{span}\{e_j : j\in J\}}}=\overline{\operatorname{span}\{e_j : j\in J\}}.$$

The latter yields (i), since we obviously have $\overline{\operatorname{span}\{e_j : j\in J\}}\subset H$. □

Proof of Lemma 2.6 By Lemma 2.5, H is a well defined set of functions. Let us show that the norm is well defined. To this end, we pick a $(b_i)\in\ell_2(\mu^{-1})$ such that $f:=\sum_{i\in I}b_ie_i$ satisfies $f=0$, where the series converges pointwise absolutely. In addition, $(b_i)\in\ell_2(\mu^{-1})\subset\ell_2(I)$ and the fact that $([e_i]_\sim)_{i\in I}$ is an ONS in $L_2(\nu)$ show that there exists a function $g:X\to\mathbb{R}$ with $[g]_\sim=\sum_{i\in I}b_i[e_i]_\sim$, where the convergence is in $L_2(\nu)$, and hence there exists a subsequence of the series that converges for ν-almost all $x\in X$ to $g(x)$. Combining both, we see $f(x)=g(x)$ for ν-almost all $x\in X$, and hence we conclude that $\sum_{i\in I}b_i[e_i]_\sim=0$ in $L_2(\nu)$. Since $([e_i]_\sim)$ is an ONS in $L_2(\nu)$, we then find $b_i=0$ for all $i\in I$, and hence every $f\in H$ has a unique representation with respect to (e_i). Consequently, the norm is well defined.

Let us now show that H is an RKHS. It is a simple routine to check that H is a Hilbert function space on X. Let us now fix an $f=\sum_{i\in I}b_ie_i\in H$, where $(b_i)\in\ell_2(\mu^{-1})$. For $x\in X$, we then have

$$\begin{aligned}\big|f(x)\big|&=\Big|\sum_{i\in I}b_ie_i(x)\Big|=\Big|\sum_{i\in I}\mu_i^{-1/2}b_i\mu_i^{1/2}e_i(x)\Big|\le\big\|(b_i)\big\|_{\ell_2(\mu^{-1})}\Big(\sum_{i\in I}\mu_ie_i^2(x)\Big)^{1/2}\\&=\Big(\sum_{i\in I}\mu_ie_i^2(x)\Big)^{1/2}\|f\|_H,\end{aligned}$$

and hence the evaluation functionals on H are continuous by (8). In other words, H is an RKHS. Moreover, by the definition of H and its norm, it is not hard to see that $(\mu_i^{1/2}e_i)_{i\in I}$ is an ONB of H, and hence H is separable. Then, by Theorem 6.2 and the fact that convergence in H implies pointwise convergence, we further conclude that (11) defines the kernel of H. In addition, $\|[e_i]_\sim\|_{L_2(\nu)}=1$ for all $i\in I$ shows that $(\|[\mu_i^{1/2}e_i]_\sim\|_{L_2(\nu)})_{i\in I}$ converges to 0, and hence H is compactly embedded into $L_2(\nu)$ by Lemma 2.4. □

Proof of Corollary 2.10 The implications (v) $\Rightarrow$ (iv), (iii) $\Rightarrow$ (i), and (ii) $\Rightarrow$ (i) are trivial, and the implication (iv) $\Rightarrow$ (i) follows from the fact that convergence in H implies pointwise convergence. Analogously, if (v) is satisfied, then $k(x,x') = \sum_{i\in I}\mu_i e_i(x)e_i(x')$ converges unconditionally for all $x,x'\in X$, and since unconditional and absolute convergence are equivalent in $\mathbb{R}$, we have shown (v) $\Rightarrow$ (ii). To show (iv) $\Rightarrow$ (iii), we first observe that we have already seen the pointwise convergence. Moreover, for a sequence $(f_n)_{n\ge 1}$ in H with $\|f_n\|_H \to 0$, the reproducing property yields

$$\bigl|f_n(x')\bigr| = \bigl|\langle f_n, k(\,\cdot\,,x')\rangle_H\bigr| \le \|f_n\|_H \,\bigl\|k(\,\cdot\,,x')\bigr\|_H = \|f_n\|_H \sqrt{k(x',x')},$$

and from this, it is easy to conclude that $f_n(x') \to 0$ uniformly in x' on every $A\subset X$ for which $\sup_{x''\in A} k(x'',x'') < \infty$. In other words, convergence in H implies uniform convergence on such A, and hence (iii) follows. Consequently, it remains to show (i) $\Rightarrow$ (v). To this end, let $\bar H$ be the RKHS induced by $(e_i)_{i\in I}$ and $(\mu_i)_{i\in I}$, and $\bar k$ be its kernel. Then (11) together with (i) shows $\bar k = k$, and since every kernel has exactly one RKHS, we find $\bar H = H$ with the same norms. Now Lemma 2.6 shows that $(\sqrt{\mu_i}e_i)$ is an ONB of $\bar H$, and hence it is an ONB of H. Theorem 6.2 then yields the assertion. □

Proof of Theorem 2.11 (i) Let us denote the operator defined by the right-hand side of (12) by $\bar S$. Since (μ_i) converges to zero, we may assume without loss of generality that $\mu_i \le 1$ for all $i\in I$. Then we have $\mu_i^{1/2} \le 1$ for all $i \in I$, and thus we find $(\mu_i^{1/2}\langle f,[e_i]_\sim\rangle_{L_2(\nu)}) \in \ell_2(I)$ for all $f\in L_2(\nu)$. Since Lemma 2.6 showed that $(\mu_i^{1/2}e_i)_{i\in I}$ is an ONB of H, we conclude that $\bar S f\in H$ and

$$\|\bar S f\|_H = \bigl\|\bigl(\mu_i^{1/2}\langle f,[e_i]_\sim\rangle_{L_2(\nu)}\bigr)\bigr\|_{\ell_2(I)} \le \bigl\|\bigl(\langle f,[e_i]_\sim\rangle_{L_2(\nu)}\bigr)\bigr\|_{\ell_2(I)} \le \|f\|_{L_2(\nu)}.$$

Consequently, $\bar S: L_2(\nu)\to H$ is well defined and continuous. Let us now fix an $f\in L_2(\nu)$ and an $h\in H$. Then we have

$$\langle \bar S f, h\rangle_H = \Bigl\langle \sum_{i\in I}\mu_i\langle f,[e_i]_\sim\rangle_{L_2(\nu)} e_i, h\Bigr\rangle_H = \sum_{i\in I}\mu_i\langle f,[e_i]_\sim\rangle_{L_2(\nu)}\langle h,e_i\rangle_H. \tag{43}$$

In addition, we have seen in Lemma 2.6 that $I_k: H\to L_2(\nu)$ defined by $h\mapsto [h]_\sim$ is well defined and continuous. Since $(\mu_i^{1/2}e_i)_{i\in I}$ is an ONB of H, we thus find

$$\begin{aligned}\langle f, I_k h\rangle_{L_2(\nu)} &= \Bigl\langle f, I_k\Bigl(\sum_{i\in I}\langle h,\mu_i^{1/2}e_i\rangle_H\,\mu_i^{1/2}e_i\Bigr)\Bigr\rangle_{L_2(\nu)} = \Bigl\langle f, \sum_{i\in I}\langle h,e_i\rangle_H\,\mu_i I_k e_i\Bigr\rangle_{L_2(\nu)}\\ &= \sum_{i\in I}\mu_i\langle f,[e_i]_\sim\rangle_{L_2(\nu)}\langle h,e_i\rangle_H,\end{aligned}$$

and hence we conclude from (43) that $\bar S = I_k^* = S_k$. Moreover, the spectral representation of T_k immediately follows from that of S_k and the continuity of I_k. The assertion on $\{\mu_i : i\in I\}$ and $([e_i]_\sim)$ is a direct consequence of this representation.

(ii) The asserted isometric isomorphism is a simple consequence of $(\mu_i^{1/2}e_i)_{i\in I}$ being an ONB of H.

(iii) Follows from integrating (11) for $x = x'$ and $\|[\tilde{e}_i]_\sim\|_{L_2(\nu)}$. □

Proof of Lemma 2.12 We have already seen in Lemma 2.2 that T_k is self-adjoint, positive, and compact. Consequently, the Spectral Theorem A.1 shows that there exist an at most countable $(\mu_i)_{i\in I} \subset (0,\infty)$ family converging to 0 with $\mu_1 \geq \mu_2 \geq \cdots > 0$ and a family $(\tilde{e}_i)_{i\in I}$ of functions such that $([\tilde{e}_i]_\sim)_{i\in I}$ is an ONS of $L_2(\nu)$ and

$$T_k f = \sum_{i\in I} \mu_i \langle f, [\tilde{e}_i]_\sim \rangle_{L_2(\nu)} [\tilde{e}_i]_\sim, \qquad f \in L_2(\nu).$$

Moreover, $(\mu_i)_{i\in I} \subset (0,\infty)$ consists of the nonzero eigenvalues (with geometric multiplicities) of T_k, and $[\tilde{e}_i]_\sim$ is an eigenvector of T_k with respect to the eigenvalue μ_i. In addition, the Spectral Theorem A.1 shows that for every ONS consisting of the eigenvectors of T_k (with geometric multiplicities), the spectral representation of T_k holds. Let us thus assume that $([\tilde{e}_i]_\sim)_{i\in I}$ is an arbitrary choice of an ONS for which the spectral representation of T_k holds. We write $e_i := \mu_i^{-1} S_k[\tilde{e}_i]_\sim \in H$ for $i \in I$. Then the diagram in Lemma 2.2 shows $[e_i]_\sim = \mu_i^{-1} T_k[\tilde{e}_i]_\sim = [\tilde{e}_i]_\sim$, and hence $([e_i]_\sim)_{i\in I}$ is an ONS of $L_2(\nu)$ consisting of the eigenvectors of T_k that belong to the nonzero eigenvalues. Moreover, the series representation (14) of T_k holds. In addition, the definition of the e_i's yields

$$\mu_i e_i = S_k[\tilde{e}_i]_\sim = S_k[e_i]_\sim$$

for all $i \in I$; i.e., we have shown (15). The last equality further shows

$$\mu_i\mu_j\langle e_i, e_j\rangle_H = \langle S_k[e_i]_\sim, S_k[e_j]_\sim\rangle_H = \langle [e_i]_\sim, T_k[e_j]_\sim\rangle_{L_2(\nu)} = \mu_j\langle [e_i]_\sim, [e_j]_\sim\rangle_{L_2(\nu)},$$

and consequently, $(\sqrt{\mu_i}e_i)_{i\in I}$ is an ONS in H. Furthermore, part (ii) of Theorem A.3 shows $\ker S_k = \ker T_k$; i.e., we have shown (16), and by considering $f_i := \sqrt{\mu_i}e_i$, part (v) of Theorem A.3 yields (17) and (18). Finally, (19) and (20) immediately follow from Theorem A.2. □

6.2 Proofs Related to Mercer's Theorem on General Domains

Proof of Theorem 3.1 To show the first set of equivalences, we recall that T_k is compact and that Lemma 2.12 yields the spectral representation of T_k for the ONS $([e_i]_\sim)_{i\in I}$. Consequently, (54) in the Spectral Theorem A.1 gives

$$L_2(\nu) = \ker T_k \oplus \overline{\operatorname{span}\{[e_i]_\sim : i \in I\}},$$

and hence $([e_i]_\sim)_{i\in I}$ is an ONB of $L_2(\nu)$ if and only if T_k is injective. Moreover, (16) in Lemma 2.12 shows that T_k is injective if and only if S_k is injective, and Theorem A.2 shows that S_k is injective if and only if S_k^* has a dense image.

To show the second series of equivalences, we define the covariance operator $C_k : H \to H$ by $C_k := S_k S_k^*$. Then (iii) of Theorem A.3 together with (14) gives the spectral representation of C_k in terms of the ONS $(f_i)_{i\in I}$ defined by $f_i := \mu_i^{-1/2} S_k[e_i]_\sim$, $i \in I$, and (54) shows

$$H = \ker C_k \oplus \overline{\operatorname{span}\{f_i : i \in I\}}.$$

With the help of (ii) of Theorem A.3 we then conclude that S_k^* is injective if and only if $(f_i)_{i\in I}$ is an ONB of H. Furthermore, Lemma 2.12 shows $f_i = \mu_i^{-1/2} S_k[e_i]_\sim = \mu_i^{1/2} e_i$, which completes the proof of (i) $\Leftrightarrow$ (ii). The equivalence (i) $\Leftrightarrow$ (iii) follows from Theorem 6.2 and Corollary 2.10. $\square$

Proof of Corollary 3.2 Let us write $H_\nu^\perp := \ker S_k^*$, where the notation is inspired by results presented later in Theorem 3.3. Since $H_\nu^\perp$ is a closed subspace of H, it is a Hilbert space that consists of functions mapping from X to $\mathbb{R}$, and for the same reason, the evaluation functionals on $H_\nu^\perp$ are continuous. Therefore, $H_\nu^\perp$ is an RKHS. In addition, $(\bar{e}_j)_{j\in J}$ considered in (22) is an ONB of $H_\nu^\perp$, and hence

$$k_\nu^\perp(x, x') := \sum_{j\in J} \bar{e}_j(x)\bar{e}_j(x'),$$

where the series converges absolutely for all $x, x' \in X$, defines the kernel of $H_\nu^\perp$ by Theorem [36, Theorem 4.20]. For $x \in X$, we then have $k_\nu^\perp(x, \cdot\,) \in H_\nu^\perp = \ker S_k^* = \ker I_k$ for all $x \in X$, and hence we conclude that $[k_\nu^\perp(x, \cdot)]_\sim = 0$. From this, the first assertion follows by (22). In addition, it follows that

$$\int_X \left|k_\nu^\perp(x, x')\right| d\nu(x') = 0, \qquad x \in X, \tag{44}$$

and by integrating over x with respect to ν, we obtain $[k_\nu^\perp]_{L_2(\nu\otimes\nu)} = 0$. Another application of (22) then yields the second assertion. Finally, if H is separable, the index set J is at most countable. Since $[\bar{e}_j]_\sim = 0$ for all $j \in J$, there then exists a ν-zero set $N \subset X$ such that $\bar{e}_j(x) = 0$ for all $x \in X \setminus N$ and *all* $j \in J$. Now the last assertion follows by yet another application of (22). $\square$

Proof of Theorem 3.3 Since H_ν is a closed subspace of H, it is a Hilbert space that consists of functions mapping from X to $\mathbb{R}$. For the same reason, the evaluation functionals on H_ν are continuous, and hence H_ν is an RKHS. In addition, (17) shows that $(\sqrt{\mu_i} e_i)_{i\in I}$ is an ONB of H_ν, and hence [36, Theorem 4.20] yields (24). The separability of H_ν is a direct consequence of the fact that I is at most countable. Moreover, (25) follows from (22) and (44). Furthermore, $H_\nu^\perp = \ker S_k^*$ yields $[H]_\sim = S_k^* H = S_k^* H_\nu = [H_\nu]_\sim$, and the last equivalence is a direct consequence of Theorem 3.1. $\square$

Proof of Corollary 3.4 Since H_ν is a subspace of H, we have $S_{k_\nu}^* h = [h]_\sim = S_k^* h$ for all $h \in H_\nu$; i.e., we have shown (27). Moreover, for fixed $x \in X$, the argument used in the proof of (7) leads to

$$I_{k_\nu}^* f(x) = \langle I_{k_\nu}^* f, k_\nu(\,\cdot\,, x)\rangle_{H_\nu} = \int_X k_\nu(x, x') f(x')\, d\nu(x')$$
$$= \int_X k(x, x') f(x')\, d\nu(x') = I_k^* f(x),$$

where in the second to last step we used (25). From this we easily obtain (26). Finally, (28) is a direct consequence of (26), (27), and Lemma 2.2. □

Proof of Corollary 3.5 Let us first show that in both cases H consists of continuous functions. To prove this in the first case, we fix an $f \in H$. Since k is continuous, [36, Lemma 4.29] shows that the canonical feature map $\Phi : X \to H$ defined by $\Phi(x) := k(\,\cdot\,, x)$, $x \in X$, is continuous, and hence the function $f = \langle f, \Phi(\,\cdot\,)\rangle_H$ is continuous. In the second case, [36, Lemma 4.28] shows that every $f \in H$ is continuous.

Since H consists of continuous functions, we now see that $N_j := \{x \in X : \bar{e}_j(x) \neq 0\}$ is an open set for all $j \in J$, where $(\bar{e}_j)_{j\in J}$ is the ONS in (22). Moreover, the discussion following (22) showed $\nu(N_j) = 0$, from which we conclude $N_j \subset X \setminus \operatorname{supp} \nu$. Consequently, $N := \bigcup_{j\in J} N_j$ satisfies $N \subset X \setminus \operatorname{supp} \nu$; i.e., for all $x \in \operatorname{supp} \nu$ and all $j \in J$, we have $\bar{e}_j(x) = 0$. From this and (22), we conclude that, for every $x \in \operatorname{supp} \nu$,

$$k(\,\cdot\,, x) = \sum_{i\in I} \mu_i e_i(x) e_i \tag{45}$$

converges unconditionally in H. By the continuity of the evaluation functionals over H, we thus find that

$$k(x', x) = \sum_{i\in I} \mu_i e_i(x) e_i(x')$$

converges unconditionally in $\mathbb{R}$ for all $x \in \operatorname{supp} \nu$ and $x' \in X$. Since in $\mathbb{R}$, unconditional convergence is equivalent to absolute convergence, we see that the latter series representation also converges absolutely; i.e., we have shown the first assertion. The uniform convergence in x' on A follows from (45) and

$$|f(x')| \leq \|k(\,\cdot\,, x')\|_H \|f\|_H = \sqrt{k(x', x')}\, \|f\|_H \leq \sup_{x''\in A} \sqrt{k(x'', x'')}\, \|f\|_H,$$

for all $x' \in A$, $f \in H$, since the latter shows that convergence in H implies uniform convergence on A. To show the last assertion, we may assume without loss of generality that $I = \mathbb{N}$. For $n \geq 1$, we define $f_n : A \to \mathbb{R}$ by

$$f_n(x) := \sum_{i=1}^{n} \mu_i e_i^2(x), \quad x \in A.$$

Since H consists of continuous functions, the functions f_n are continuous. Moreover, the function $f : A \to \mathbb{R}$ defined by $f(x) := k(x, x)$, $x \in A$, is continuous since k is continuous. Now our previous considerations showed $f_n(x) \to f(x)$ for all $x \in A$, and by the form of f_n, this convergence is also monotone. Dini's theorem then yields $\sup_{x\in A}|f_n(x)-f(x)|\to 0$. Let us now consider the kernels $k_n : A\times A\to\mathbb{R}$ defined by

$$k_n(x,x') := \sum_{i=1}^{n} \mu_i e_i(x) e_i(x'), \quad x,x'\in A.$$

For $x,x'\in A$, our previous considerations on the pointwise convergence together with the Cauchy-Schwarz inequality yields

$$\begin{aligned}\left|k_n(x,x')-k(x,x')\right| &= \left|\sum_{i=n+1}^{\infty} \mu_i e_i(x) e_i(x')\right| \le \sqrt{\sum_{i=n+1}^{\infty} \mu_i e_i^2(x)} \cdot \sqrt{\sum_{i=n+1}^{\infty} \mu_i e_i^2(x')} \\ &= \sqrt{f_n(x)-f(x)}\cdot\sqrt{f_n(x')-f(x')}.\end{aligned}$$

Using $\sup_{x\in A}|f_n(x)-f(x)|\to 0$, we then obtain the uniform convergence on $A\times A$. □

Proof of Corollary 3.6 We have seen in the proof of Corollary 3.5 that H consists of continuous functions. Given an $f\in H$ with $f\neq 0$, the set $\{f\neq 0\}$ is thus an open and nonempty subset of X, and hence we have $\nu(\{f\neq 0\})>0$. This shows $[f]_\sim\neq 0$; i.e., $S_k^* : H\to L_2(\nu)$ is injective. Applying Theorem 3.1, we then find that $(\sqrt{\mu_i}e_i)_{i\in I}$ is an ONB of H, and since I is at most countable, the first assertion follows. To prove the second assertion, we observe that the conditions imply $\|k\|_{\mathcal{L}_2(\nu)}<\infty$, and hence Lemma 2.3 shows that H is compactly embedded into $L_2(\nu)$. Now the assertion follows from the first assertion. □

Proof of Theorem 3.7 (i) ⇒ (ii) Follows directly by considering the set $(X\backslash N)^n$.

(ii) ⇒ (iii) Follows directly by considering the case $n=1$.

(iii) ⇒ (i) Let $(\mu_i)_{i\in I}$ and $(e_i)_{i\in I}\subset H$ be the families obtained by Lemma 2.12. Furthermore, let $(\bar{e}_j)_{j\in J}$ be an ONS in H such that the union of $(\sqrt{\mu_i}e_i)_{i\in I}$ and $(\bar{e}_j)_{j\in J}$ is an ONB of H. Then we have the representation (22), from which we conclude that

$$\sum_{i\in I}\mu_i e_i^2(x) = k_\nu(x,x) = k(x,x) = \sum_{i\in I}\mu_i e_i^2(x) + \sum_{j\in J}\bar{e}_j^2(x), \quad x\in X\backslash N.$$

Since $k(x,x)\in\mathbb{R}$, we consequently obtain $\bar{e}_j(x)=0$ for all $j\in J$ and $x\in X\backslash N$. For $x\in X\backslash N$, the representation (22) thus yields

$$k(\cdot,x) = \sum_{i\in I}\mu_i e_i(x) e_i + \sum_{j\in J}\bar{e}_j(x)\bar{e}_j = \sum_{i\in I}\mu_i e_i(x) e_i = k_\nu(\cdot,x).$$

From this the assertion immediately follows. □

Proof of Corollary 3.8 (ii) ⇒ (i) Follows from Theorem 3.7 and the definition of k_ν by considering $\bar{e}_i := e_i$ and $\bar{k} := k_\nu$.

(i) $\Rightarrow$ (ii) For $n=1$, the assumption on the Gram matrices yields a measurable $\bar{N} \subset X$ with $\nu(\bar{N})=0$ and $\bar{k}(x,x)=k(x,x)$ for all $x \in X \setminus \bar{N}$. For $i \in I$, we define $N_i := \{e_i \neq \bar{e}_i\}$. Then we have $\nu(N_i)=0$ for all $i \in I$, and since I is at most countable, $\tilde{N} := N \cup \bar{N} \cup \bigcup_{i \in I} N_i$ is measurable with $\nu(\tilde{N})=0$. For $x \in X \setminus \tilde{N}$, we then find

$$k(x,x)=\bar{k}(x,x)=\sum_{i \in I} \mu_i \bar{e}_i^2(x)=\sum_{i \in I} \mu_i e_i^2(x)=k_\nu(x,x).$$

This shows that $\tilde{N}$ is the desired ν-zero set. □

Proof of Theorem 3.10 By the spectral theorem, T has a representation of the form

$$Tf=\sum_{i \in I} \mu_i \langle f, [\tilde{e}_i]_\sim \rangle_{L_2(\nu)} [\tilde{e}_i]_\sim, \quad f \in L_2(\nu),$$

where $(\tilde{e}_i)_{i \in I}$ is a family of functions $\tilde{e}_i : X \to \mathbb{R}$ such that $([\tilde{e}_i]_\sim)_{i \in I}$ is an ONS in $L_2(\nu)$. Moreover, since T is positive and nuclear, the eigenvalues $(\mu_i)_{i \in I}$ of T are positive and summable, and hence we obtain

$$\int_X \sum_{i \in I} \mu_i \tilde{e}_i^2 \, d\nu = \sum_{i \in I} \mu_i \int_X \tilde{e}_i^2 \, d\nu = \sum_{i \in I} \mu_i < \infty,$$

by Beppo Levi's theorem. Consequently, there exists a measurable $N \subset X$ such that $\nu(N)=0$ and

$$\sum_{i \in I} \mu_i \tilde{e}_i^2(x) < \infty, \quad x \in X \setminus N.$$

We define $e_i := \mathbf{1}_{X \setminus N} \tilde{e}_i$. Then we have $[e_i]_\sim = [\tilde{e}_i]_\sim$ and $\sum_{i \in I} \mu_i e_i^2(x) < \infty$ for *all* $x \in X$. Let H be the RKHS induced by (e_i) and (μ_i). Clearly, H has a measurable kernel k which satisfies $\|k\|^2_{\mathcal{L}_2(\nu)} = \sum_{i \in I} \mu_i < \infty$ by Theorem 2.11, and Theorem 2.11 further shows that T_k enjoys the spectral representation

$$T_k f=\sum_{i \in I} \mu_i \langle f, [e_i]_\sim \rangle_{L_2(\nu)} [e_i]_\sim, \quad f \in L_2(\nu).$$

Since $[e_i]_\sim = [\tilde{e}_i]_\sim$ for all $i \in I$, we conclude that $T=T_k$; i.e., we have shown (31). Moreover, (32) holds by construction. Finally, using Beppo Levi's theorem, we can compute the Hilbert-Schmidt norm of S_k^* by

$$\begin{aligned}\|S_k^*\|_{\mathrm{HS}}^2 &= \sum_{i \in I} \|S_k^* \sqrt{\mu_i} e_i\|^2_{L_2(\nu)} = \sum_{i \in I} \int_X \mu_i e_i^2(x) \, d\nu(x) = \int_X \sum_{i \in I} \mu_i e_i^2(x) \, d\nu(x) \\ &= \|k\|^2_{\mathcal{L}_2(\nu)} < \infty.\end{aligned}$$

Since $\|T\|_{\mathrm{nuc}} = \|T_k\|_{\mathrm{nuc}} = \|S_k^* S_k\|_{\mathrm{nuc}} = \|S_k^*\|_{\mathrm{HS}} \|S_k\|_{\mathrm{HS}} = \|S_k^*\|^2_{\mathrm{HS}}$, we then find the last assertion. □

6.3 Proofs Related to Powers of Kernels and Integral Operators

Proof of Proposition 4.2 Clearly, we only need to prove that H_ν^β is independent of the choice of $(e_i)_{i\in I} \subset H$ in Lemma 2.12. To this end, let $J \subset I$ be an index set in which every eigenvalue appears exactly once; that is, for all $i \in I$, there exists exactly one $j \in J$ with $\mu_i = \mu_j$. For $j \in J$, we further write $I_j := \{i \in I : \mu_i = \mu_j\}$ for the finite set of indices that share their eigenvalue with j. Furthermore, let $C_k : H \to H$ be the covariance operator defined by $C_k := S_k \circ S_k^*$. Then Theorem A.3 shows that $f_i := \mu_i^{-1/2} S_k[e_i]_\sim$, $i \in I$, defines an ONS in H that consists of the eigenvectors of C_k. For $j \in J$, we write

$$E(\mu_j) := \ker(\mu_j \operatorname{id}_H - C_k)$$

for the eigenspace of μ_j. Clearly, $E(\mu_j)$ is independent of the choice of the family $(e_i)_{i\in I} \subset H$ in Lemma 2.12. Furthermore, $(f_i)_{i\in I_j}$ is a finite ONB of $E(\mu_j)$ with respect to the norm $\|\cdot\|_H$, and since (15) shows $f_i = \mu_i^{-1/2} S_k[e_i]_\sim = \mu_j^{1/2} e_i$ for all $i \in I_j$, we conclude that

$$f = \sum_{i\in I_j} \langle f, f_i\rangle_H f_i = \mu_j^{1/2} \sum_{i\in I_j} \langle f, f_i\rangle_H e_i = \mu_j^{1/2-\beta/2} \sum_{i\in I_j} \langle f, f_i\rangle_H \mu_i^{\beta/2} e_i$$

for all $f \in E(\mu_j)$. Consequently, we have $E(\mu_j) \subset H_\nu^\beta$ for all $j \in J$, and the formula further shows that the corresponding two norms of $f \in E(\mu_j)$ can be computed by

$$\|f\|_H^2 = \sum_{i\in I_j} \bigl|\langle f, f_i\rangle_H\bigr|^2,$$

$$\|f\|_{H_\nu^\beta}^2 = \mu_j^{1-\beta} \sum_{i\in I_j} \bigl|\langle f, f_i\rangle_H\bigr|^2.$$

This yields $\|f\|_{H_\nu^\beta} = \mu_j^{(\beta-1)/2} \|f\|_H$ for all $f \in E(\mu_j)$; i.e., for such f, the norm $\|f\|_{H_\nu^\beta}$ is independent of the choice of the family $(e_i)_{i\in I} \subset H$ in Lemma 2.12. Let us write $H_j := E(\mu_j)$ and equip this subspace of H_ν^β with the norm of H_ν^β. Then, by construction, H_ν^β is the direct Hilbert space sum of the family $(H_j)_{j\in J}$, and since each H_j and its norm are independent of the choice of $(e_i)_{i\in I} \subset H$ in Lemma 2.12, H_ν^β is independent of that choice, too. Finally, every RKHS has a unique kernel, and hence k_ν^β is also independent of the chosen family $(e_i)_{i\in I} \subset H$. □

Proof of Lemma 4.3 The first assertion follows from combining (13) for the induced kernel k_ν^β and the definition (33) of T_k^β. In addition, $[H_\nu^\beta]_\sim = [H]_\sim^\beta$ follows from the definitions of the two spaces and the fact that H_ν^β is compactly embedded into $L_2(\nu)$. Now let us fix an $\alpha \ge \beta$. Since (μ_i) converges to 0, we may further assume without loss of generality that $\mu_i \le 1$ for all $i \in I$. Then we have $\mu_i^\alpha \le \mu_i^\beta$ for all $i \in I$, and thus the third assertion follows. The last assertion follows from $\ell_2(\mu^{-\alpha}) \hookrightarrow \ell_2(\mu^{-\beta})$. □

Proof of Proposition 4.4 Using Beppo Levi's theorem and the fact that $([e_i]_\sim)_{i\in I}$ is an ONS in $L_2(\nu)$, we obtain

$$\int_X \sum_{i\in I} \mu_i^\beta e_i^2 \, d\mu = \sum_{i\in I} \mu_i^\beta \int_X e_i^2 \, d\mu = \sum_{i\in I} \mu_i^\beta, \tag{46}$$

where all values may be infinite.

(i) $\Rightarrow$ (ii) If $(\mu_i)_{i\in I}$ is β-summable, then (46) shows that $\sum_{i\in I} \mu_i^\beta e_i^2$ is ν-integrable and thus the sum is finite modulo a ν-zero set N. This yields (37). Moreover, the induced kernel $\bar{k}_\nu^\beta$ satisfies

$$\bar{k}_\nu^\beta(x,x) = \sum_{i\in I} \mu_i^\beta \bar{e}_i^2(x) = \sum_{i\in I} \mu_i^\beta e_i^2(x), \quad x \in X\backslash N, \tag{47}$$

and hence we obtain $\|\bar{k}_\nu^\beta\|_{L_2(\nu)}^2 = \sum_{i\in I} \mu_i^\beta$ by (46).

(ii) $\Rightarrow$ (i) The condition $\|\bar{k}_\nu^\beta\|_{L_2(\nu)} < \infty$ together with the construction of $\bar{k}_\nu^\beta$, which ensures (47), shows that the left-hand side of (46) is finite. □

Proof of Proposition 4.5 Let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as in Lemma 2.12. Moreover, let us fix an $(a_i) \in \ell_2(I)$. Then $\sum_{i\in I} a_i \mu_i^{\beta/2}[e_i]_\sim$ converges in $[H]_\sim^\beta$, and since $[H]_\sim^\beta \hookrightarrow L_\infty(\nu)$, it also converges in $L_\infty(\nu)$. Let us further consider the bounded linear map $\varphi : L_\infty(\nu) \to \mathcal{L}_\infty(X)$ defined by (58), where $\mathcal{L}_\infty(X)$ denotes the space of bounded measurable functions $f : X \to \mathbb{R}$ equipped with the usual supremum norm. Moreover, note that φ actually exists, because Theorem B.2 ensures the existence of a ν-lifting $\rho : \mathcal{L}_\infty(X) \to \mathcal{L}_\infty(X)$. Since φ is continuous, we then obtain

$$\varphi\biggl(\sum_{i\in I} a_i \mu_i^{\beta/2}[e_i]_\sim\biggr) = \sum_{i\in I} a_i \mu_i^{\beta/2} \varphi\bigl([e_i]_\sim\bigr),$$

where the latter convergence is in $\mathcal{L}_\infty(X)$. Moreover, convergence in $\mathcal{L}_\infty(X)$ implies pointwise convergence. For the *function* $\bar{e}_i := \varphi([e_i]_\sim)$, we thus find that

$$\sum_{i\in I} a_i \mu_i^{\beta/2} \bar{e}_i(x)$$

converges for *all* $x \in X$. Since $(a_i) \in \ell_2(I)$ was arbitrarily chosen, Lemma 2.5 then shows $\sum_{i\in I} \mu_i^\beta \bar{e}_i^2(x) < \infty$. Now the assertion follows from (59), which implies $[\bar{e}_i]_\sim = [\varphi([e_i]_\sim)]_\sim = [e_i]_\sim$ and the fact that I is at most countable. □

For the proof of Theorem 4.6, we have to recall the following result that interpolates L_2-spaces whose underlying measures are absolutely continuous with respect to a measure ν:

Lemma 6.3 *Let ν be a measure on a measurable space Ω and $w_0 : \Omega \to [0,\infty)$ and $w_1 : \Omega \to [0,\infty)$ be measurable functions. For $0 < \beta < 1$, we define $w_\beta := w_0^{1-\beta} w_1^\beta$. Then we have*

$$\bigl[L_2(w_0\, d\nu), L_2(w_1\, d\nu)\bigr]_{\beta,2} = L_2(w_\beta\, d\nu),$$

and the norms of these two spaces are equivalent. Moreover, this result still holds for weights $w_0 : \Omega \to (0, \infty)$ *and* $w_1 : \Omega \to [0, \infty]$ *if one uses the convention* $0 \cdot \infty := 0$ *in the definition of the weighted spaces.*

Proof The first assertion is proven in, e.g., [41, Lemma 23.1] and the second assertion can be shown by carefully repeating the proof for [41, Lemma 23.1]. □

To motivate another result we need for the proof of Theorem 4.6, we assume for a moment that (34) holds; that is, the power k_ν^β exists. Then we obtain

$$\operatorname{ran} T_k^{\beta/2} = \operatorname{ran} T_{k_\nu^\beta}^{1/2} = \operatorname{ran} S_{k_\nu^\beta}^* S_{k_\nu^\beta}^{1/2} = \left[H_\nu^\beta\right]_\sim, \tag{48}$$

where the first equality follows from Lemma 4.3 and the last two follow from Theorem 2.11 and $S_{k_\nu^\beta}^* = [\,\cdot\,]_\sim$. In other words, the range of fractional powers can be identified with powers of H when considered as subsets of $L_2(\nu)$. As we will see later, this observation will be a key step in the proof of Theorem 4.6. Unfortunately, however, we do not have the means to ensure that (34) actually holds. To address this issue, the following lemma shows the formally weaker but conceptionally similar equality $\operatorname{ran} T_k^{\beta/2} = [H]_\sim^\beta$:

Lemma 6.4 *Let X be a measurable space, ν be a measure on X, and k be a measurable kernel on X whose RKHS H is compactly embedded into $L_2(\nu)$. Let $(e_i)_{i\in I} \subset H$ and $(\mu_i)_{i\in I} \subset (0,\infty)$ be as in Lemma 2.12 and, for some $\beta > 0$, we define $[H]_\sim^\beta$ by (36). Then, for all $f \in L_2(\nu)$, we have $T_k^\beta f \in [H]_\sim^\beta$, with*

$$\left\| T_k^\beta f \right\|_{[H]_\sim^\beta} \le \mu_1^{\beta/2} \|f\|_{L_2(\nu)}.$$

Consequently, $R_\beta : L_2(\nu) \to [H]_\sim^\beta$ defined by $R_\beta f := T_k^\beta f$ for all $f \in L_2(\nu)$ is well defined and continuous. Moreover, its adjoint satisfies $R_\beta^* h = h$ for all $h \in [H]_\sim^\beta$, and hence we have

$$\begin{tikzcd} L_2(\nu) \arrow[rr, "T_k^\beta"] \arrow[dr, "R_\beta"'] & & L_2(\nu) \\ & {[H]_\sim^\beta} \arrow[ur, "R_\beta^*"'] & \end{tikzcd}$$

Furthermore, $f \mapsto T_k^{\beta/2} f$ defines an isometric isomorphism between $[H]_\sim^0 = \overline{\operatorname{span}\{[e_i]_\sim : i \in I\}}$ and $[H]_\sim^\beta$, and thus we have

$$\operatorname{ran} T_k^{\beta/2} = [H]_\sim^\beta.$$

Finally, we have $[H]_\sim^1 = [H]_\sim$, and if (34) holds, we further have $[H]_\sim^\beta = [H_\nu^\beta]_\sim$.

Proof Let us fix an $f \in L_2(\nu)$. Since $(\ker T_k)^\perp = \overline{\operatorname{span}\{[e_i]_\sim : i \in I\}} = [H]^0_\sim$ by Theorem A.1, there exist an $f_1 \in \ker T_k = \ker T_k^\beta$ and an $f_2 \in [H]^0_\sim$ with $f = f_1 + f_2$ and $f_1 \perp f_2$. Moreover, we have

$$\big\| \big(\langle f_2, [e_i]_\sim\rangle_{L_2(\nu)}\big)\big\|^2_{\ell_2(I)} = \|f_2\|^2_{L_2(\nu)} \le \|f\|^2_{L_2(\nu)}$$

by Parseval's identity and $f_1 \perp f_2$, and hence we obtain

$$\|T_k^\beta f\|^2_{[H]^\beta_\sim} = \sum_{i\in I} \frac{\mu_i^{2\beta} |\langle f_2, [e_i]_\sim\rangle_{L_2(\nu)}|^2}{\mu_i^\beta} \le \mu_1^\beta \sum_{i\in I} \big|\langle f_2, [e_i]_\sim\rangle_{L_2(\nu)}\big|^2 \le \mu_1^\beta \|f\|^2_{L_2(\nu)}.$$

Let us now show that $R^*_\beta h = h$ for all $h \in [H]^\beta_\sim$. To this end, we fix an $f \in L_2(\nu)$ and an $h \in [H]^\beta_\sim$ with $h = \sum_{i\in I} b_i [e_i]_\sim$. Let us split f into $f = f_1 + f_2$, where f_1 and f_2 are as above. Since $\ker R_\beta = \ker T_k$, we then find $R_\beta f_1 = 0$. In addition, we have $f_1 \perp [e_i]_\sim$ for all $i \in I$, and hence we obtain $h \perp f_1$ in $L_2(\nu)$. The definition of the norm $\|\cdot\|_{[H]^\beta_\sim}$ and the corresponding inner product thus yields

$$\begin{aligned}\langle h, R_\beta f\rangle_{[H]^\beta_\sim} &= \langle h, R_\beta f_2\rangle_{[H]^\beta_\sim} = \sum_{i\in I} \frac{b_i \mu_i^\beta \langle f_2, [e_i]_\sim\rangle_{L_2(\nu)}}{\mu_i^\beta} = \sum_{i\in I} b_i \big\langle f_2, [e_i]_\sim\big\rangle_{L_2(\nu)} \\ &= \langle h, f_2\rangle_{L_2(\nu)} = \langle h, f\rangle_{L_2(\nu)};\end{aligned}$$

i.e., we have shown $R^*_\beta h = h$. Now, the diagram easily follows. Let us finally show the claimed isometric isomorphism. To this end, we first observe that $([e_i]_\sim)_{i\in I}$ is an ONB of $[H]^0_\sim$ and $\|f\|_{[H]^0_\sim} = \|f\|_{L_2(\nu)}$ for all $f \in [H]^0_\sim$. Therefore, Parseval's identity yields

$$\|T_k^{\beta/2} f\|^2_{[H]^\beta_\sim} = \sum_{i\in I} \frac{\mu_i^\beta |\langle f, [e_i]_\sim\rangle_{L_2(\nu)}|^2}{\mu_i^\beta} = \|f\|^2_{L_2(\nu)} = \|f\|^2_{[H]^0_\sim}, \quad f \in [H]^0_\sim;$$

i.e., $f \mapsto T_k^{\beta/2} f$ is an isometric injection from $[H]^0_\sim$ to $[H]^\beta_\sim$. To show the surjectivity, we fix an $h \in [H]^\beta_\sim$ with $h = \sum_{i\in I} b_i [e_i]_\sim$. Then we have $(b_i)_{i\in I} \in \ell_2(\mu^{-\beta})$, and for $(a_i)_{i\in I}$ defined by $a_i := \mu_i^{-\beta/2} b_i$, we thus obtain $(a_i)_{i\in I} \in \ell_2$. Consequently, we can define $f := \sum_{i\in I} a_i [e_i]_\sim \in [H]^0_\sim$. An easy calculation then shows

$$T_k^{\beta/2} f = \sum_{i\in I} \mu_i^{\beta/2} a_i [e_i]_\sim = \sum_{i\in I} b_i [e_i]_\sim = h;$$

i.e., we have shown the surjectivity. The final assertions follow easily. □

Proof of Theorem 4.6 We have already seen in Lemma 6.4 that $\operatorname{ran} T_k^{\beta/2} = [H]^\beta_\sim$. For the first assertion it thus suffices to show that $[L_2(\nu), [H]_\sim]_{\beta,2} = [H]^\beta_\sim$. To this end, we first observe that $(\sqrt{\mu_i}[e_i]_\sim)_{i\in I}$ is an ONB of $[H]^1_\sim = [H]_\sim$ by the definition of $[H]_\sim$ and its norm. Let us fix an ONB $(\tilde e_j)_{j\in J}$ of $\ker T_k$. Then the union of $([e_i]_\sim)_{i\in I}$ and $(\tilde e_j)_{j\in J}$ forms an ONB of $L_2(\nu)$. Without loss of generality, we may further assume that $I \cap J = \emptyset$, which will help to avoid unpleasant technicalities. Let us now fix an $f \in L_2(\nu)$. We write

$$a_i := \begin{cases} \langle f, [e_i]_\sim \rangle_{L_2(\nu)} & \text{if } i \in I, \\ \langle f, \tilde{e}_i \rangle_{L_2(\nu)} & \text{if } i \in J; \end{cases}$$

that is, $(a_i)_{i \in I \cup J}$ is the family of Fourier coefficients of f with respect to the ONB above. Let us further write $\mu_i := 0$ for all $i \in J$ and $\mu := (\mu_i)_{i \in I \cup J}$. Because of the convention $0/0 := 0 \cdot \infty := 0$, the identification of $h = \sum_{i \in I} b_i [e_i]_\sim \in [H]_\sim$ with the family $(b_i)_{i \in I \cup J} \in \ell_2(\mu^{-1})$ that satisfies $b_i = 0$ for all $i \in J$ then gives a one-to-one relation between $[H]_\sim$ and $\ell_2(\mu^{-1})$. Moreover, this identification is isometric, since

$$\|h\|^2_{[H]_\sim} = \Big\| \sum_{i \in I} b_i [e_i]_\sim \Big\|^2_{[H]^1_\sim} = \sum_{i \in I \cup J} \frac{b_i^2}{\mu_i} = \big\| (b_i) \big\|^2_{\ell_2(\mu^{-1})}.$$

In addition, this identification together with Parseval's identity yields

$$\|f - h\|^2_{L_2(\nu)} = \sum_{i \in I \cup J} |a_i - b_i|^2 = \big\| (a_i - b_i) \big\|^2_{\ell_2(I \cup J)}.$$

For $t > 0$, we consequently find

$$\begin{aligned} K\big(f, t, L_2(\nu), [H]_\sim\big) &= \inf_{h \in [H]_\sim} \|f - h\|_{L_2(\nu)} + t\|h\|_{[H]_\sim} \\ &= \inf_{(b_i) \in \ell_2(\mu^{-1})} \big\| (a_i - b_i) \big\|_{\ell_2(I \cup J)} + t \big\| (b_i) \big\|_{\ell_2(\mu^{-1})} \\ &= K\big(a, t, \ell_2(I \cup J), \ell_2\big(\mu^{-1}\big)\big). \end{aligned}$$

From this we immediately find the equivalence

$$f \in \big[L_2(\nu), [H]_\sim\big]_{\beta,2} \quad \Longleftrightarrow \quad (a_i) \in \big[\ell_2(I \cup J), \ell_2\big(\mu^{-1}\big)\big]_{\beta,2},$$

with $\|f\|_{[L_2(\nu),[H]_\sim]_{\beta,2}} = \|(a_i)\|_{[\ell_2(I \cup J),\ell_2(\mu^{-1})]_{\beta,2}}$. Moreover, the second part of Lemma 6.3 applied to the counting measure on $I \cup J$ yields

$$\big[\ell_2(I \cup J), \ell_2\big(\mu^{-1}\big)\big]_{\beta,2} = \ell_2\big(\mu^{-\beta}\big).$$

Using the definition of $[H]^\beta_\sim$ in (36), we hence obtain $[L_2(\nu), [H]_\sim]_{\beta,2} = [H]^\beta_\sim$. To show the second assertion, we fix an $h \in [H]^\beta_\sim$. We define

$$b_i := \begin{cases} \langle h, [e_i]_\sim \rangle_{L_2(\nu)} & \text{if } i \in I, \\ 0 & \text{if } i \in J, \end{cases}$$

which, by the definition of powers, implies $(b_i) \in \ell_2(\mu^{-\beta})$ with $\|h\|_{[H]^\beta_\sim} = \|(b_i)\|_{\ell_2(\mu^{-\beta})}$. Moreover, Lemma 6.3 shows that there exist constants $c_\beta, C_\beta \in (0, \infty)$ independent of (b_i) such that

$$c_\beta \big\| (b_i) \big\|_{\ell_2(\mu^{-\beta})} \le \big\| (b_i) \big\|_{[\ell_2(I \cup J),\ell_2(\mu^{-1})]_{\beta,2}} \le C_\beta \big\| (b_i) \big\|_{\ell_2(\mu^{-\beta})}.$$

In addition, our considerations on the K-functional in the first part of the proof showed

$$\big\|(b_i)\big\|_{[\ell_2(I\cup J),\ell_2(\mu^{-1})]_{\beta,2}} = \|h\|_{[L_2(\nu),[H]_\sim]_{\beta,2}},$$

and consequently, the norms $\|\cdot\|_{[H]_\sim^\beta}$ and $\|\cdot\|_{[L_2(\nu),[H]_\sim]_{\beta,2}}$ are equivalent on $[H]_\sim^\beta = [L_2(\nu),[H]_\sim]_{\beta,2}$. Finally, we have seen in Lemma 6.4 that $f \mapsto T_k^{\beta/2} f$ is an isometric isomorphism between $\overline{\operatorname{span}\{[e_i]_\sim : i\in I\}}$ and $[H]_\sim^\beta$, and hence the last assertion follows. □

For the proof of Corollary 4.7, we need the following lemma that recalls the ordering of interpolation spaces:

Lemma 6.5 *Let E and F be Banach spaces. Then the following statements are true*:

(i) *For all* $0<\beta<1$ *and* $1\le r\le r'\le\infty$, *we have* $[E,F]_{\beta,r}\hookrightarrow [E,F]_{\beta,r'}$.
(ii) *If* $F\hookrightarrow E$, *then, for all* $0<\beta<\beta'<1$, *we have* $[E,F]_{\beta',\infty}\hookrightarrow [E,F]_{\beta,1}$.

Proof The proof of the first assertion can be found in, e.g., [6, Proposition 1.10 on p. 301]. To show the second assertion, we may assume without loss of generality that $\|\cdot\|_E\le\|\cdot\|_F$. Let us then fix an $x\in[E,F]_{\beta',\infty}$ with $\|x\|_{\beta',\infty}\le 1$; that is, we have

$$K(x,t)\le t^{\beta'},\qquad t>0. \tag{49}$$

Moreover, for all $y\in F$, we have $\|x\|_E\le\|x-y\|_E+\|y\|_E\le\|x-y\|_E+\|y\|_F$, and hence we find

$$\|x\|_E\le K(x,1)\le 1.$$

From this we conclude

$$K(x,t)\le\|x-0\|_E+t\|0\|_F=\|x\|_E\le 1,\qquad t>0. \tag{50}$$

Combining (49) and (50), we now obtain

$$\|x\|_{\beta,1}=\int_0^\infty t^{-\beta-1}K(x,t)\,dt\le\int_0^1 t^{\beta'-\beta-1}\,dt+\int_1^\infty t^{-\beta-1}\,dt=\frac{1}{\beta'-\beta}+\frac{1}{\beta};$$

i.e. we have shown the second assertion. □

Proof of Corollary 4.7 This is an immediate consequence of Theorem 4.6 and Lemma 6.5. □

6.4 Proofs Related to Bounded Powers of Kernels

Proof of Lemma 5.1 (ii) ⇒ (i) Trivial.

(iii) ⇒ (ii) For $f\in H$ and $x\in X\setminus N$, the reproducing property yields

$$\big|f(x)\big|=\big|\langle f,k(\cdot,x)\rangle_H\big|\le\|f\|_H\,\big\|k(\cdot,x)\big\|_H=\|f\|_H\sqrt{k(x,x)}\le\|f\|_H\,\kappa.$$

(i) ⇒ (iii) We write $\kappa := \|[\,\cdot\,]_\sim : H \to L_\infty(\nu)\|$. Let us first consider the case of kernels satisfying the assumptions of Corollary 3.5. For $x \in X$, we further define

$$N_x := \bigl\{x' \in X : k\bigl(x', x\bigr) > \bigl\|k(\,\cdot\,, x)\bigr\|_{L_\infty(\nu)}\bigr\}.$$

Then N_x is open and satisfies $\nu(N_x) = 0$. This implies $N_x \subset X \setminus \operatorname{supp} \nu$, and hence the open set $N := \cup_{x\in X} N_x$ satisfies $N \subset X \setminus \operatorname{supp} \nu$ and thus $\nu(N) = 0$. Now, for $x \in X \setminus N$, we obtain

$$k(x, x) \le \bigl\|k(\,\cdot\,, x)\bigr\|_{L_\infty(\nu)} \le \kappa \bigl\|k(\,\cdot\,, x)\bigr\|_H = \kappa \sqrt{k(x, x)}. \tag{51}$$

Let us now consider the case where H is separable. For $f \in H$, the set

$$N_f := \bigl\{x \in X : \bigl|f(x)\bigr| > \kappa \|f\|_{L_\infty(\nu)}\bigr\}$$

is measurable with $\nu(N_f) = 0$. Since H is separable, there further exists a countable subset $D \subset H$ that is dense in H. Then $N := \cup_{f\in D} N_f$ clearly satisfies $\nu(N) = 0$. Let us now fix an $f \in H$ and an $\varepsilon > 0$. Then there exists a sequence $(f_i) \subset D$ such that $\|f_i - f\|_H \to 0$ and $\|f_i\|_H \le \|f\|_H + \varepsilon$ for all $i \ge 1$. Since convergence in H implies pointwise convergence, we further conclude that $f_i(x) \to f(x)$ for all $x \in X$. Moreover, for $x \in X \setminus N$ and $i \ge 1$, we have

$$\bigl|f_i(x)\bigr| \le \|f_i\|_{L_\infty(\nu)} \le \kappa \|f_i\|_H \le \kappa\bigl(\|f\|_H + \varepsilon\bigr).$$

Combining these observations, we find $|f(x)| \le \kappa(\|f\|_H + \varepsilon)$ for all $x \in X \setminus N$, and by letting $\varepsilon \to 0$, we find $|f(x)| \le \kappa \|f\|_H$ for $x \in X \setminus N$. Setting $f := k(\,\cdot\,, x)$, we again obtain (51). □

Proof of Theorem 5.3 (i) ⇔ (ii) Because of Proposition 4.5, we can consider the RKHS $\bar{H}_\nu^\beta$ introduced in front of Proposition 4.4. Since this RKHS is separable by Lemma 2.6 and it further satisfies $[\bar{H}_\nu^\beta]_\sim = [H]_\sim^\beta$, we then see that the equivalence is a direct consequence of Lemma 5.1 when applied to $\bar{H}_\nu^\beta$ and $\bar{k}_\nu^\beta$.

(i) ⇔ (iii) By Theorem 4.6, we know that there exist constants $c_\beta, C_\beta \in (0, \infty)$ such that

$$c_\beta \|f\|_{L_2(\nu)} \le \bigl\|T_k^{\beta/2} f\bigr\|_{[H]_\sim^\beta} \le C_\beta \|f\|_{L_2(\nu)}$$

for all $f \in [H]_\sim^0$, and since $\ker T_k^{\beta/2} = \ker T_k = ([H]_\sim^0)^\perp$ by Theorem A.1, it is easy to see that the second inequality actually holds for all $f \in L_2(\nu)$. In addition, $[H]_\sim^\beta \hookrightarrow L_\infty(\nu)$ is equivalent to the existence of a constant $C \in [0, \infty)$ such that $\|h\|_{L_\infty(\nu)} \le C \|h\|_{[H]_\sim^\beta}$ for all $h \in [H]_\sim^\beta = \operatorname{ran} T_k^{\beta/2}$. For $h = T_k^{\beta/2} f$, we then obtain the assertion.

To show the additional assertions, we first assume that (i) is true. Then we have $[L_2(\nu), [H]_\sim]_{\beta,1} \hookrightarrow [L_2(\nu), [H]_\sim]_{\beta,2} = [H]_\sim^\beta \hookrightarrow L_\infty(\nu)$ by Lemma 6.5 and Theorem 4.6. In addition, [6, Proposition 2.10 on p. 316] shows that $[L_2(\nu), [H]_\sim]_{\beta,1} \hookrightarrow L_\infty(\nu)$ is equivalent to (39).

Let us now assume that ν is finite. Then (ii) implies $\sum_{i\in I} \mu_i^\beta < \infty$ by Proposition 4.4. Finally, we have already mentioned that [6, Proposition 2.10 on p. 316] shows that (39) implies $[L_2(\nu), [H]_\sim]_{\beta,1} \hookrightarrow L_\infty(\nu)$. Consequently, Theorem 4.6 and a double application of Lemma 6.5 imply that

$$\begin{aligned}[H]_\sim^{\beta+\varepsilon} &= \big[L_2(\nu), [H]_\sim\big]_{\beta+\varepsilon,2} \hookrightarrow \big[L_2(\nu), [H]_\sim\big]_{\beta+\varepsilon,\infty} \\ &\hookrightarrow \big[L_2(\nu), [H]_\sim\big]_{\beta,1} \hookrightarrow L_\infty(\nu),\end{aligned}$$

and we have already seen that $[H]_\sim^{\beta+\varepsilon} \hookrightarrow L_\infty(\nu)$ implies $\sum_{i\in I} \mu_i^{\beta+\varepsilon} < \infty$. □

Proof of Theorem 5.4 As in the proof of Theorem 4.6, we fix an ONB $(\tilde{e}_j)_{j\in J}$ of $\ker T_k$ with $I \cap J = \emptyset$. In addition, we write

$$a_i := \begin{cases} \langle f^*_{L,\mathrm{P}}, [e_i]_\sim\rangle_{L_2(\nu)} & \text{if } i \in I, \\ \langle f^*_{L,\mathrm{P}}, \tilde{e}_i\rangle_{L_2(\nu)} & \text{if } i \in J, \end{cases}$$

for the Fourier coefficients of $f^*_{L,\mathrm{P}}$. For an arbitrary $h = \sum_{i\in I} b_i[e_i]_\sim \in [H]_\sim$, Parseval's identity then yields

$$\big\| f^*_{L,\mathrm{P}} - h\big\|^2_{L_2(\nu)} = \sum_{i\in I} (a_i - b_i)^2 + \sum_{i\in J} |a_i|^2$$

and

$$\|h\|^2_{[H]_\sim} = \|h\|^2_{[H]^1_\sim} = \sum_{i\in I} \mu_i^{-1} b_i^2 .$$

Consequently, we have

$$\lambda\|h\|^2_{[H]_\sim} + \big\| f^*_{L,\mathrm{P}} - h\big\|^2_{L_2(\nu)} = \sum_{i\in I} \big(\lambda\mu_i^{-1} b_i^2 + (a_i - b_i)^2\big) + \sum_{i\in J} |a_i|^2, \tag{52}$$

and minimizing the left-hand side over $h \in [H]_\sim$ amounts to the same as minimizing the first sum on the right-hand side over $(b_i)_{i\in I} \in \ell_2(\mu^{-1})$, where $\mu := (\mu_i)_{i\in I}$. Now, for each fixed $i \in I$, the function

$$b_i \mapsto \lambda\mu_i^{-1} b_i^2 + (a_i - b_i)^2$$

is minimized by $b_i^* := \frac{\mu_i}{\lambda+\mu_i} a_i$. Moreover, this definition yields

$$\begin{aligned}\big\|\big(b_i^*\big)\big\|^2_{\ell_2(\mu^{-1})} &= \sum_{i\in I} \frac{(b_i^*)^2}{\mu_i} = \sum_{i\in I} \frac{\mu_i}{(\lambda+\mu_i)^2} a_i^2 \le \sup_{i\in I} \frac{\mu_i}{(\lambda+\mu_i)^2} \sum_{i\in I} a_i^2 \\ &\le \frac{1}{4\lambda} \sum_{i\in I} a_i^2 < \infty,\end{aligned}$$

where in the second to last step we used that the function $t \mapsto \frac{t}{(\lambda+t)^2}$ attains its maximum at $t^* := \lambda$. From this, we conclude that (b_i^*) is the unique minimizer over $\ell_2(\mu^{-1})$ of the right-hand side of (52). On the other hand, it is well known that, for every measurable function $f : X \to \mathbb{R}$, we have $\mathcal{R}_{L,\mathrm{P}}(f) = \|[f]_\sim - f^*_{L,\mathrm{P}}\|^2_{L_2(\nu)}$. In particular, if we split $f_{\mathrm{P},\lambda}$ into $f_{\mathrm{P},\lambda} = f_1 + f_2$ for suitable $f_1 \in H_\nu$ and $f_2 \in H_\nu^\perp = \ker S_k^*$, we have $[f_2]_\sim = 0$, and hence we obtain $\mathcal{R}_{L,\mathrm{P}}(f_{\mathrm{P},\lambda}) = \mathcal{R}_{L,\mathrm{P}}(f_1)$. Since this yields

$$\lambda\|f_1\|_H^2 + \mathcal{R}_{L,\mathrm{P}}(f_1) \le \lambda\|f_{\mathrm{P},\lambda}\|_H^2 + \mathcal{R}_{L,\mathrm{P}}(f_{\mathrm{P},\lambda})$$

and $f_{\mathrm{P},\lambda}$ is the unique minimizer of the regularized risk, we conclude that $f_{\mathrm{P},\lambda} = f_1$; that is, $f_{\mathrm{P},\lambda} \in H_\nu$. Using the isometric identification of $[H]_\sim = [H_\nu]_\sim$ with H_ν via $S^*_{k_\nu}$, see (29), we then find that $[f_{\mathrm{P},\lambda}]_\sim$ is the unique minimizer over $[H]_\sim$ of the left-hand side of (52). Combining this with our considerations on the minimizer of the right-hand side of (52), we conclude that

$$[f_{\mathrm{P},\lambda}]_\sim = \sum_{i\in I} b_i^*[e_i]_\sim = \sum_{i\in I} \mu_i^{-1/2} b_i^* \sqrt{\mu_i}[e_i]_\sim .$$

Since $(\sqrt{\mu_i}[e_i]_\sim)_{i\in I}$ is an ONB of $[H]_\sim$, we thus find

$$\big\langle [f_{\mathrm{P},\lambda}]_\sim, \sqrt{\mu_i}[e_i]_\sim \big\rangle_{[H]_\sim} = \mu_i^{-1/2} b_i^* = \frac{\mu_i^{1/2}}{\lambda+\mu_i} a_i .$$

Finally, using the isometric identification of $[H]_\sim = [H_\nu]_\sim$ with H_ν via $S^*_{k_\nu}$ another time, we obtain the assertion. □

Proof of Corollary 5.5 Let us write $a_i := \langle f^*_{L,\mathrm{P}}, [e_i]_\sim\rangle_{L_2(\nu)}$ for $i \in I$. Then we know by Theorem 4.6 and the definition of $[H]^\beta_\sim$ in (36) that

$$\sum_{i\in I} \frac{a_i^2}{\mu_i^\beta} < C^2,$$

where $C \in [0,\infty)$ is a suitable constant. For $x \in X$, we thus obtain by Theorem 5.4

$$\begin{aligned}\big|f_{\mathrm{P},\lambda}(x)\big| &= \Big|\sum_{i\in I}\big\langle f_{\mathrm{P},\lambda}, \sqrt{\mu_i} e_i\big\rangle_H \sqrt{\mu_i} e_i(x)\Big| \le \sum_{i\in I} \frac{\mu_i}{\lambda+\mu_i}\big|a_i e_i(x)\big| \\ &\le \sup_{i\ge 1} \frac{\mu_i^{1+\beta/2-\alpha/2}}{\lambda+\mu_i} \cdot \sum_{i\in I}\big|\mu_i^{-\beta/2} a_i\big| \cdot \big|\mu_i^{\alpha/2} e_i(x)\big| .\end{aligned}$$

Let us now fix the ν-zero set N and the constant κ in part (ii) of Theorem 5.3. By Hölder's inequality and the previous estimates, we then find

$$\big|f_{\mathrm{P},\lambda}(x)\big| \le \kappa\, C \sup_{t>0} \frac{t^{1+\beta/2-\alpha/2}}{\lambda+t}$$

for all $x \in X\backslash N$. A simple optimization with respect to t then yields the assertion. □

Appendix A: Related Operators and the Spectral Theorem

This appendix recalls some facts from the spectral theory of compact, self-adjoint operators acting between Hilbert spaces. We begin with the classical spectral theorem, see, e.g., [18, Theorem V.2.10 on p. 260] or [45, Theorem VI.3.2].

Theorem A.1 (Spectral Theorem) *Let H be a Hilbert space and $A : H \to H$ be a compact, positive, and self-adjoint operator. Then there exist an at most countable ONS $(e_i)_{i\in I}$ of H and a family $(\mu_i)_{i\in I}$ converging to 0 such that $\mu_1 \geq \mu_2 \geq \cdots > 0$ and*

$$Ax = \sum_{i\in I} \mu_i \langle x, e_i\rangle e_i, \quad x \in H, \tag{53}$$

$$H = \ker A \oplus \overline{\operatorname{span}\{e_i : i \in I\}}. \tag{54}$$

Moreover, $(\mu_i)_{i\in I}$ is the family of nonzero eigenvalues of A (including geometric multiplicities), and, for all $i \in I$, e_i is an eigenvector for μ_i. Finally, both (53) and (54) actually hold for all ONSs $(\tilde{e}_i)_{i\in I}$ of H for which, for all $i \in I$, the vector $\tilde{e}_i$ is an eigenvector of μ_i.

The next well known theorem, see, e.g., [27, Theorem 12.10], relates the image of an operator B to the null-space of its adjoint B^*.

Theorem A.2 *Let H_1 and H_2 be Hilbert spaces and $B : H_1 \to H_2$ be a bounded linear operator. Then we have*

$$\overline{\operatorname{ran} B} = \big(\ker B^*\big)^{\perp}.$$

In particular, B^* is injective if and only if B has a dense image.

The following theorem lists other well-known facts about the eigenvalues and the spectral representations of certain operators. These facts are widely known, but since we are unaware of a reference for the particular formulation we need, we decided to include its relatively straightforward proof for the sake of completeness. The name for this theorem was borrowed from [23, Sect. 3.3.4].

Theorem A.3 (Principle of Related Operators) *Let H_1 and H_2 be Hilbert spaces and $B : H_1 \to H_2$ be a bounded linear operator. We define the self-adjoint and positive operators $A_1 : H_1 \to H_1$ and $A_2 : H_2 \to H_2$ by $A_1 := B^*B$ and $A_2 := BB^*$, respectively. Then the following statements are true:*

(i) *Given a $\mu > 0$, we denote the eigenspaces of A_1 and A_2 that correspond to μ by*

$$E_1(\mu) := \ker(\mu \operatorname{id}_{H_1} - A_1),$$

$$E_2(\mu) := \ker(\mu \operatorname{id}_{H_2} - A_2).$$

Then the map

$$
\begin{aligned}
B_\mu : E_1(\mu) &\to E_2(\mu),\\
x &\mapsto Bx
\end{aligned}
$$

is well defined, bijective, and, in addition, we have $B_\mu^* B_\mu = \mu\, \mathrm{id}_{E_1(\mu)}$.

(ii) *We have* $\ker A_1 = \ker B$ *and* $\ker A_2 = \ker B^*$.

(iii) *Assume that* A_1 *is compact, and let* $(e_i)_{i\in I}$ *be an at most countable ONS of* H_1 *and* $(\mu_i)_{i\in I}$ *be a family converging to* 0 *such that* $\mu_1 \geq \mu_2 \geq \cdots > 0$ *and*

$$
A_1 x = \sum_{i\in I} \mu_i \langle x, e_i\rangle_{H_1} e_i, \qquad x \in H_1,
$$

where we note that there exist such $(e_i)_{i\in I}$ *and* $(\mu_i)_{i\in I}$ *by Theorem A.1. For* $i \in I$, *we define* $f_i := \mu_i^{-1/2} Be_i$. *Then* $(f_i)_{i\in I}$ *is an ONS of* H_2, *and* A_2 *has the spectral representation*

$$
A_2 y = \sum_{i\in I} \mu_i \langle y, f_i\rangle_{H_2} f_i, \qquad y \in H_2. \tag{55}
$$

Finally, we have $e_i = \mu_i^{-1/2} B^* f_i$ *for all* $i \in I$.

(iv) A_1 *is compact if and only if* A_2 *is compact.*

(v) *If* A_1 *is compact, we have* $\overline{\operatorname{ran} B} = \overline{\operatorname{span}\{f_i : i \in I\}} = \overline{\operatorname{ran} A_2}$ *and* $\overline{\operatorname{ran} B^*} = \overline{\operatorname{span}\{e_i : i \in I\}} = \overline{\operatorname{ran} A_1}$.

Proof (i) It is easy to check that A_1 and A_2 are indeed self-adjoint and positive. Let us first show that B_μ is well defined, that is, that $Bx \in E_2(\mu)$ for all $x \in E_1(\mu)$. To this end, we pick an $x \in E_1(\mu)$; i.e., we have $A_1 x = \mu x$. From this we conclude $A_2 Bx = BB^* Bx = BA_1 x = \mu Bx$, and thus we have $Bx \in E_2(\mu)$. Similarly, for $x \in E_1(\mu)$ with $Bx = 0$, we obtain $\mu x = Ax = B^* Bx = 0$, and since $\mu > 0$, we conclude $x = 0$; i.e., B_μ is injective. By interchanging the role of B and B^*, we analogously see that

$$
\begin{aligned}
B_\mu^* : E_2(\mu) &\to E_1(\mu),\\
x &\mapsto B^* x
\end{aligned}
$$

is well defined and injective. To show that B_μ is surjective, we now pick an $y \in E_2(\mu)$ and define $x := \mu^{-1} B^* y$. Our previous consideration then gives $x \in E_1(\mu)$, and since we further have $Bx = \mu^{-1} BB^* y = \mu^{-1} A_2 y = y$, we obtain the surjectivity of B_μ. The last assertion is a consequence of $B^* B = A_1$.

(ii) By symmetry, it suffices to show $\ker A_1 = \ker B$. Moreover, the inclusion $\ker A_1 \supset \ker B$ immediately follows from $A_1 = B^* B$. To show the converse inclusion, we fix an $x \in \ker A_1$. Then the definition of A_1 yields $0 = \langle A_1 x, x\rangle_{H_1} = \langle Bx, Bx\rangle_{H_2}$, which implies $x \in \ker B$.

(iii) Let us show that $(f_i)_{i\in I}$ is an ONS of H_2. To this end, we fix $i, j \in I$ and find

$$
\langle f_i, f_j\rangle_{H_2} = \frac{1}{\sqrt{\mu_i \mu_j}} \langle Be_i, Be_j\rangle_{H_2} = \frac{1}{\sqrt{\mu_i \mu_j}} \langle e_i, A_1 e_j\rangle_{H_1} = \sqrt{\frac{\mu_j}{\mu_i}} \langle e_i, e_j\rangle_{H_1},
$$

and since $(e_i)_{i\in I}$ is an ONS, we then conclude that $(f_i)_{i\in I}$ is an ONS. Moreover, by (i) we have

$$A_2 f_i = BB^* f_i = \mu_i^{-1/2} BB^* B e_i = \mu_i^{1/2} B e_i = \mu_i f_i, \quad i \in I. \tag{56}$$

By Theorem A.1 and (ii), the compactness of A_1 further implies

$$\begin{aligned}\overline{\operatorname{ran} B} &= \overline{B\big(\ker A_1 \oplus \overline{\operatorname{span}\{e_i : i\in I\}}\big)} = \overline{B\big(\ker B \oplus \overline{\operatorname{span}\{e_i : i\in I\}}\big)} \\ &= \overline{B\big(\overline{\operatorname{span}\{e_i : i\in I\}}\big)} \\ &= \overline{\operatorname{span}\{Be_i : i\in I\}} \\ &= \overline{\operatorname{span}\{f_i : i\in I\}}.\end{aligned}$$

Using Theorem A.2 and (ii), we then find

$$H_2 = \ker B^* \oplus \overline{\operatorname{ran} B} = \ker A_2 \oplus \overline{\operatorname{span}\{f_i : i\in I\}},$$

and combining this with (56), we obtain (55) by Theorem A.1. The last assertion follows from $\mu_i^{-1/2} B^* f_i = \mu_i^{-1} B^* B e_i = e_i$, where in the last step we used (i).

(iv) If A_1 is compact, we obtain the spectral representation (55) for A_2, and hence A_2 is compact. The inverse implication follows by symmetry.

(v) The proof of (iii) has already shown $\overline{\operatorname{ran} B} = \overline{\operatorname{span}\{f_i : i\in I\}} = \overline{\operatorname{ran} A_2}$, and the second equality follows by symmetry and (iv). □

Appendix B: Liftings

In this appendix, we briefly recall some results related to liftings on the space of bounded measurable functions. To this end, we assume that $(X, \mathcal{A})$ is a measurable space. We denote by

$$\mathcal{L}_\infty(X) := \{f : X \to \mathbb{R} \mid f \text{ bounded and measurable}\}$$

the space of all bounded measurable functions on X and equip this space with the usual supremum norm $\|\cdot\|_\infty$. Furthermore, if ν is a σ-finite measure[2] on $(X, \mathcal{A})$, we define, for a measurable $f : X \to \mathbb{R}$,

$$\|f\|_{\mathcal{L}_\infty(\nu)} := \inf\big\{a \ge 0 : \big\{x \in X : \big|f(x)\big| > a\big\} \text{ is a } \nu\text{-zero set}\big\}.$$

This leads to the space

$$\mathcal{L}_\infty(\nu) := \big\{f : X \to \mathbb{R} \mid f \text{ measurable and } \|f\|_{\mathcal{L}_\infty(\nu)} < \infty\big\}$$

[2]For the sake of simplicity, we restrict our considerations to σ-finite measures, since otherwise we would have to deal with local ν-zero sets and, later, when dealing with liftings, with technically involved assumptions on ν. Since for the applications we are most interested in we typically have a probability measure, the σ-finiteness is no restriction.

of *essentially bounded*, measurable functions on X. By considering a sequence $\alpha_n \searrow \|f\|_{\mathcal{L}_\infty(\nu)}$, it is straightforward to show that the infimum in the definition of $\|\cdot\|_{\mathcal{L}_\infty(\nu)}$ is actually attained; that is,

$$\nu\big(\{|f| > \|f\|_{\mathcal{L}_\infty(\nu)}\}\big) = 0, \qquad f \in \mathcal{L}_\infty(\nu). \tag{57}$$

Furthermore, we write $L_\infty(\nu) := \mathcal{L}_\infty(\nu)_{/\sim}$ for the quotient space with respect to the usual equivalence relation $f \sim g :\Leftrightarrow \nu(\{f \neq g\}) = 0$ on $\mathcal{L}_\infty(\nu)$. From (57), we can then conclude that the linear map

$$\begin{aligned} I_\nu : \mathcal{L}_\infty(X) &\to L_\infty(\nu), \\ f &\mapsto [f]_\sim \end{aligned}$$

is a metric surjection, and hence its quotient map $\bar{I}_\nu : \mathcal{L}_\infty(X)_{/\sim} \to L_\infty(\nu)$, which is given by $\bar{I}([f]_\sim) = [f]_\sim$, is an isometric isomorphism. Here we note that, given an $f \in \mathcal{L}_\infty(X)$, the equivalence class $[f]_\sim$ in $L_\infty(\nu)$ consists, in general, of more functions than the corresponding equivalence class $[f]_\sim$ in $\mathcal{L}_\infty(X)_{/\sim}$. Nevertheless, $\bar{I}_\nu$ gives us a canonical tool to identify the spaces $\mathcal{L}_\infty(X)_{/\sim}$ and $L_\infty(\nu)$. Our next goal is to find a continuous and linear right-inverse of I_ν. To this end, we recall the definition of a lifting on $\mathcal{L}_\infty(X)$.

Definition B.1 Let $(X, \mathcal{A})$ be a measurable space and ν be a σ-finite measure on $(X, \mathcal{A})$. We say that a map $\rho : \mathcal{L}_\infty(X) \to \mathcal{L}_\infty(X)$ is a ν-lifting if the following conditions are satisfied:

(i) ρ is an algebra homeomorphism; that is, ρ is linear and $\rho(fg) = \rho(f)\rho(g)$ for all $f, g \in \mathcal{L}_\infty(X)$.
(ii) $\rho(\mathbf{1}_X) = \mathbf{1}_X$.
(iii) $[\rho(f)]_\sim = [f]_\sim$ for all $f \in \mathcal{L}_\infty(X)$.
(iv) $\rho(f) = \rho(g)$ for all $f, g \in \mathcal{L}_\infty(X)$ with $[f]_\sim = [g]_\sim$.

If $\rho : \mathcal{L}_\infty(X) \to \mathcal{L}_\infty(X)$ is a ν-lifting and $f \in \mathcal{L}_\infty(X)$ satisfies $f(x) \geq 0$, then $h := \sqrt{f} \in \mathcal{L}_\infty(X)$, and hence we obtain

$$\rho(f) = \rho\big(h^2\big) = \rho^2(h) \geq 0.$$

Consequently, ρ also respects the ordering on $\mathcal{L}_\infty(X)$. From this we can conclude that ρ is also continuous with respect to $\|\cdot\|_\infty$. Indeed, if we have an $f \in \mathcal{L}_\infty(X)$ with $\|f\|_\infty \leq 1$, we find $-\mathbf{1}_X \leq f \leq \mathbf{1}_X$, and since ρ is linear and respects the ordering, we obtain

$$-\mathbf{1}_X = \rho(-\mathbf{1}_X) \leq \rho(f) \leq \rho(\mathbf{1}_X) = \mathbf{1}_X;$$

that is, $\|\rho(f)\|_\infty \leq 1$. In other words, we have shown $\|\rho\| \leq 1$. Moreover, from conditions (iii) and (iv), we can conclude that

$$\ker \rho = \big\{f \in \mathcal{L}_\infty(X) : \nu\big(\{f \neq 0\}\big) = 0\big\} = [0]_\sim,$$

where the equivalence class $[0]_\sim$ is the one in $\mathcal{L}_\infty(X)$. Consequently, the quotient map $\bar{\rho} : \mathcal{L}_\infty(X)_{/\sim} \to \mathcal{L}_\infty(X)$, defined by $[f]_\sim \mapsto f$, is a well-defined, bounded, linear, and injective operator with $\|\bar{\rho}\| \le 1$. Moreover, from condition (ii), we obtain

$$\big[\bar{\rho}\big([f]_\sim\big)\big]_\sim = \big[\rho(f)\big]_\sim = [f]_\sim, \quad f \in \mathcal{L}_\infty(X);$$

that is, $[\,\cdot\,]_\sim \circ \bar{\rho} = \mathrm{id}_{\mathcal{L}_\infty(X)_{/\sim}}$. Considering the isometric isomorphism $\bar{I}_\nu^{-1} : L_\infty(\nu) \to \mathcal{L}_\infty(X)_{/\sim}$, we can now define the bounded, linear, and injective operator

$$\varphi := \bar{\rho} \circ \bar{I}_\nu^{-1} : L_\infty(\nu) \to \mathcal{L}_\infty(X), \tag{58}$$

which, by construction and our previous considerations, satisfies $\|\varphi\| \le 1$ and

$$\big[\varphi\big([f]_\sim\big)\big]_\sim = [f]_\sim, \quad f \in \mathcal{L}_\infty(\nu), \tag{59}$$

where the outer $[\,\cdot\,]_\sim$ on the left-hand side and $[\,\cdot\,]_\sim$ on the right-hand side both refer to equivalence classes in $\mathcal{L}_\infty(X)_{/\sim}$. Applying $\bar{I}_\nu$ to both sides, we conclude that $I_\nu \circ \varphi = \mathrm{id}_{L_\infty(\nu)}$; i.e., φ is the desired right-inverse of I_ν. In other words, φ picks from each equivalence class of $L_\infty(\nu)$ a bounded representative in a linear and continuous way. Moreover, it is not hard to check that φ is actually an algebra homeomorphism with $\varphi([\mathbf{1}_X]_\sim) = \mathbf{1}_X$, and, in addition, it also respects the ordering of $L_\infty(\nu)$ and $\mathcal{L}_\infty(X)$. Consequently, the map φ has highly desirable properties. So far, however, we have not shown that there actually exists a ν-lifting, and hence we do not know whether a map φ with the above properties exists. This gap is filled by the following theorem:

Theorem B.2 (Existence of liftings) *Let $(X, \mathcal{A})$ be a measurable space and ν be a σ-finite measure on $(X, \mathcal{A})$ such that $\mathcal{A}$ is ν-complete. Then there exists a ν-lifting $\rho : \mathcal{L}_\infty(X) \to \mathcal{L}_\infty(X)$.*

Proof See [39, Theorem 3.2], where we note that σ-finite measures are strictly localizable and Carathéodory completeness of $\mathcal{A}$ equals ν-completeness. □

As [39, Theorem 3.2] shows, the σ-finiteness of the measure is not necessary if one assumes, instead, some other, technically more involved conditions on ν. Finally, further information on liftings and their applications can be found in, e.g., [24, 42].

References

1. Agler, J., McCarthy, J.E.: Pick Interpolation and Hilbert Function Spaces. Am. Math. Soc., Providence (2002)
2. Ali, S.T., Antoine, J.-P., Gazeau, J.-P.: Coherent States, Wavelets and Their Generalizations. Springer, New York (2000)
3. Alpay, D. (ed.): Reproducing Kernel Spaces and Applications. Birkhäuser Verlag, Basel (2003)
4. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
5. Bauer, H.: Measure and Integration Theory. De Gruyter, Berlin (2001)
6. Bennett, C., Sharpley, R.: Interpolation of Operators. Academic Press, Boston (1988)

7. Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics. Kluwer, Boston (2004)
8. Caponnetto, A., De Vito, E.: Fast rates for regularized least-squares algorithm. Technical Report CBCL Paper #248, AI Memo #2005-013, MIT, Cambridge, MA (2005)
9. Caponnetto, A., De Vito, E.: Optimal rates for regularized least squares algorithm. Found. Comput. Math. **7**, 331–368 (2007)
10. Cesa-Bianchi, N., Lugosi, G.: Prediction, Learning, and Games. Cambridge University Press, Cambridge (2006)
11. Conway, J.B.: A Course in Functional Analysis, 2nd edn. Springer, New York (1990)
12. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. Cambridge University Press, Cambridge (2000)
13. Cucker, F., Smale, S.: On the mathematical foundations of learning. Bull. Am. Math. Soc. **39**, 1–49 (2002)
14. Cucker, F., Zhou, D.X.: Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, Cambridge (2007)
15. De Vito, E., Caponnetto, A., Rosasco, L.: Model selection for regularized least-squares algorithm in learning theory. Found. Comput. Math. **5**, 59–85 (2005)
16. Hein, M., Bousquet, O.: Kernels, associated structures and generalizations. Technical Report 127, Max-Planck-Institute for Biological Cybernetics (2004)
17. Hille, E.: Introduction to general theory of reproducing kernels. Rocky Mt. J. Math. **2**, 321–368 (1972)
18. Kato, T.: Perturbation Theory for Linear Operators, 2nd edn. Springer, Berlin–New York (1976)
19. König, H.: Eigenvalue Distribution of Compact Operators. Birkhäuser, Basel (1986)
20. Mendelson, S., Neeman, J.: Regularization in kernel learning. Ann. Stat. **38**, 526–565 (2010)
21. Meschkowski, H.: Hilbertsche Räume mit Kernfunktion. Springer, Berlin (1962)
22. Novak, E., Woźniakowski, H.: Tractability of Multivariate Problems. Linear Information, vol. 1. European Mathematical Society (EMS), Zürich (2008)
23. Pietsch, A.: Eigenvalues and s-Numbers. Geest & Portig K.-G., Leipzig (1987)
24. Rao, M.M.: Measure Theory and Integration, 2nd edn. Dekker, New York (2004)
25. Riesz, F., Nagy, B.Sz.: Functional Analysis, 2nd edn. Dover, New York (1990)
26. Ritter, K.: Average-Case Analysis of Numerical Problems. Lecture Notes in Math., vol. 1733. Springer, Berlin (2000)
27. Rudin, W.: Functional Analysis, 2nd edn. McGraw-Hill, New York (1991)
28. Saitoh, S.: Theory of Reproducing Kernels and Applications. Longman Scientific & Technical, Harlow (1988)
29. Saitoh, S.: Integral Transforms, Reproducing Kernels and Their Applications. Longman Scientific & Technical, Harlow (1997)
30. Saitoh, S., Alpay, D., Ball, J.A., Ohsawa, T. (eds.): Reproducing Kernels and Their Applications, Kluwer Academic, Dordrecht (1999)
31. Schölkopf, B., Smola, A.J.: Learning with Kernels. MIT Press, Cambridge (2002)
32. Schölkopf, B., Smola, A.J., Müller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**, 1299–1319 (1998)
33. Shawe-Taylor, J., Williams, C.K.I., Cristianini, N., Kandola, J.: On the eigenspectrum of the Gram matrix and the generalization error of kernel-PCA. IEEE Trans. Inf. Theory **51**, 2510–2522 (2005)
34. Smale, S., Zhou, D.-X.: Estimating the approximation error in learning theory. Anal. Appl. **1**, 17–41 (2003)
35. Smale, S., Zhou, D.-X.: Learning theory estimates via integral operators and their approximations. Constr. Approx. **26**, 153–172 (2007)
36. Steinwart, I., Christmann, A.: Support Vector Machines. Springer, New York (2008)
37. Steinwart, I., Hush, D., Scovel, C.: Optimal rates for regularized least squares regression. In: Dasgupta, S., Klivans, A. (eds.) Proceedings of the 22nd Annual Conference on Learning Theory, pp. 79–93 (2009)
38. Steinwart, I., Hush, D., Scovel, C.: Training SVMs without offset. J. Mach. Learn. Res. **12**, 141–202 (2011)
39. Strauss, W., Macheras, N.D., Musiał, K.: Liftings. In: Pap, E. (ed.) Handbook of Measure Theory, vol. II, pp. 1131–1184. Elsevier, Amsterdam (2002)
40. Sun, H.: Mercer theorem for RKHS on noncompact sets. J. Complex. **21**, 337–349 (2005)
41. Tartar, L.: An Introduction to Sobolev Spaces and Interpolation Spaces. Springer, Berlin (2007)
42. Ionescu Tulcea, A., Ionescu Tulcea, C.: Topics in the Theory of Lifting. Springer, New York (1969)

43. Wahba, G.: Spline Models for Observational Data. Series in Applied Mathematics, vol. 59. SIAM, Philadelphia (1990)
44. Wendland, H.: Scattered Data Approximation. Cambridge University Press, Cambridge (2005)
45. Werner, D.: Funktionalanalysis. Springer, Berlin (1995)
46. Yao, Y., Rosasco, L., Caponnetto, A.: On early stopping in gradient descent learning. Constr. Approx. **26**, 289–315 (2007)
47. Zhou, D.-X.: The covering number in learning theory. J. Complex. **18**, 739–767 (2002)