Ann. Funct. Anal. 4 (2013), no. 1, 64–88
ANNALS OF FUNCTIONAL ANALYSIS
ISSN: 2008-8752 (electronic)
URL:www.emis.de/journals/AFA/

POSITIVE DEFINITENESS, REPRODUCING KERNEL HILBERT SPACES AND BEYOND

J. C. FERREIRA[1*] AND V. A. MENEGATTO[2]

Communicated by V. Valov

ABSTRACT. Positive definiteness, reproducing kernel Hilbert spaces, integral operators and Mercer's theorem in its various formats are common topics in many branches of mathematics. In this paper we review and upgrade upon some recent results that always involve at least two of them and indicate a few directions in which additional research could be carried out.

1. INTRODUCTION

This paper is mainly concerned with a few concepts with we immediately recall. Let X be a non-empty set and K a *positive definite kernel* on X, that is, a function $K : X \times X \to \mathbb{C}$ satisfying the inequality

$$\sum_{i,j=1}^{n} \overline{c_i} c_j K(x_i, x_j) \geq 0,$$

whenever $n \geq 1$, $\{x_1, x_2, \ldots, x_n\}$ is a subset of X and $\{c_1, c_2, \ldots, c_n\}$ is a subset of $\mathbb{C}$. For $x \in X$, let us write K^x to denote the function $y \in X \to K(y, x) \in \mathbb{C}$. The unique Hilbert space $\mathcal{H}_K$ defined by the properties
- $K^x \in \mathcal{H}_K$, $x \in X$;
- the linear span of $\{K^x : x \in X\}$ is dense in $\mathcal{H}_K$;
- the inner product $\langle \cdot, \cdot \rangle_K$ of $\mathcal{H}_K$ satisfies $f(x) = \langle f, K^x \rangle_K$, $x \in X$, $f \in \mathcal{H}_K$,

is called the reproducing kernel Hilbert space (RKHS) associated to K. The equality above is usually referred to as the *reproducing property* and usually plays

Date: Received: 11 September 2012; Accepted: 8 October 2012.
* Corresponding author.
2010 *Mathematics Subject Classification.* Primary 47Bxx; Secondary 46Cxx, 42A82, 45C05, 45P05, 46E22.

Key words and phrases. Positive definite kernel, reproducing kernel Hilbert space, Mercer's theorem, integral operator.

a significant role in most problems where RKHSs enter. For instance, they enter in the solution of many problems in Approximation Theory, Learning Theory, Functional Analysis and many other areas as one can ratify in the references [1, 9, 11, 26, 27, 28, 29]. If we assume X is endowed with a convenient measure ν and the integral operator $\mathcal{K} : L^2(X,\nu) \to L^2(X,\nu)$ given by the formula

$$\mathcal{K}(f)(x) = \int_X K(x,y)f(y)\,d\nu(y), \quad f \in L^2(X,\nu), \quad x \in X,$$

is well defined then it is quite frequent to find results dealing with the analysis of many questions related to K and $\mathcal{K}$ in connection with the space $\mathcal{H}_K$. We mention Saitoh's book [26] for theoretical aspects of reproducing kernel Hilbert spaces and the papers [9, 31, 33] for applications. Other references included there increment the list of sources for the exigent reader.

One of our intentions in writing this paper is then to describe some of the problems we alluded to in the previous paragraph, but putting them in a slightly more general setting not considered before. A second intention is to justify properly all the results announced in [12]. These two goals will be achieved in Sections 1-5. Subsequently, after restricting ourselves to the case in which X is an open subset of Euclidean space, we consider some results involving the differentiability of positive definite kernels. In the cases in which a certain degree of smoothness is guaranteed, we tackle on Mercer's representation for certain derivatives of the kernels. This topic is of interest in the search for improved decay rates for eigenvalues and singular values of integral operators, when an additional assumption on certain derivatives of the generating kernel is to be used. These results and some circumstances where they were used before are described in Section 6. There we also consider the search for reproducing like formulas for derivatives of functions in the associated RKHS. Finally, still considering X as a subset of an Euclidean space, Section 7 is devoted to the description of some results involving both positive definite kernels and the usual Fourier transform in Euclidean space, a topic not fully studied yet.

Despite being a bit disconnected, the paper is intended to serve as motivation for future research related to the different topics covered here or elsewhere.

2. Embedding the RKHS in other spaces

A good starting point for what we intend to cover in this section is reference [31]. There, some interesting properties involving $\mathcal{H}_K$ and spectral properties of $\mathcal{K}$ related to Mercer's theorem were obtained in the case in which X is a metric space endowed with a strictly positive measure ν and having a compactness structure of the following type: $X = \cup_{n=1}^{\infty} X_n$, where each X_n is a compact set of finite measure, $X_n \subset X_{n+1}$, $n = 1, 2, \ldots$, and every compact subset of X is a subset of some X_n. Assuming a technical assumption on the operator $\mathcal{K}$, a more general version of the classical Mercer's theorem was established. Among properties related to $\mathcal{H}_K$, it can be found proved there a intimate connection between the inner product of $\mathcal{H}_K$ and the inner product of $L^2(X,\nu)$, which allows one to see $\mathcal{H}_K$ as a subspace of $L^2(X,\nu)$. This study continued in [33], now assuming that X was just a metric space endowed with a probability measure ν. A particular result obtained in this setting was the compactness of the inclusion of $\mathcal{H}_K$ into spaces of continuous functions, under two basic assumptions: compactness of X and continuity of K. The proofs of these results used the metric or compactness assumptions on the space in a decisive manner.

Here, we tackle some of these very same issues when the space X has neither a metric nor a compactness structure. Precisely, we will consider cases in which X is a topological space endowed with a (complete) Borel measure on X. Thus, throughout this section and the other two to come, X and ν are as above and the kernel K is positive definite on X, carrying no additional assumptions.

As continuity of K is not assumed, we begin with a little discussion about it. Usually, the continuity of K is closely related to the continuity of the so-called *feature map* $\eta : X \to \mathcal{H}_K$ given by

$$\eta(x) = K^x, \quad x \in X.$$

Since

$$\|\eta(x)\|_K^2 = K(x,x) := \kappa(x), \quad x \in X,$$

η is uniformly bounded if and only if κ is bounded. Also, the equality

$$\|\eta(x)-\eta(y)\|_K^2 = \langle K^x - K^y, K^x - K^y\rangle_K = \kappa(x) - K(x,y) - K(y,x) + \kappa(y), \tag{2.1}$$

along with the inequality

$$\begin{aligned} |K(x,y) - K(u,v)| &= |\langle \eta(x), \eta(y)\rangle_K - \langle \eta(u), \eta(v)\rangle_K| \\ &\leq |\langle \eta(x) - \eta(u), \eta(y)\rangle_K| + |\langle \eta(u), \eta(y) - \eta(v)\rangle_K|, \end{aligned}$$

for all $x, y, u, v \in X$, imply an equivalence between the continuity of K and that of η, as we will detail in the next lemma. We write $C(X)$ to denote the set of all complex continuous functions on X.

Lemma 2.1. *The kernel K is continuous if and only if η is so. In particular, if K is continuous then $\mathcal{H}_K \subset C(X)$.*

Proof. Suppose that the function K is continuous. Let A be an open set in $\mathcal{H}_K$ and $g = \eta(x) \in A$, for some $x \in X$. Pick, as we can, an $\epsilon > 0$ such that the ball in $\mathcal{H}_K$ of radius $\sqrt{3\epsilon}$ centered in $\eta(x)$ is a subset of A and also an open set $\mathcal{O} \subset X$ for which $x \in \mathcal{O}$ and

$$|K(x,x) - K(y,z)| < \epsilon, \quad y, z \in \mathcal{O}.$$

It follows that

$$|K(y,y) - K(y,x)| \leq |K(y,y) - K(x,x)| + |K(x,x) - K(y,x)| < 2\epsilon, \quad y \in \mathcal{O}.$$

Hence, by using the comments preceding the lemma, we can see that

$$\|\eta(x) - \eta(y)\|_K^2 \leq |K(x,x) - K(x,y)| + |K(y,y) - K(y,x)| \leq 3\epsilon, \quad y \in \mathcal{O}.$$

This clearly means that $\eta(\mathcal{O}) \subset A$ and η is continuous. For the converse, we use the equality

$$K(x,y) = \langle \eta(y), \eta(x)\rangle_K, \quad x, y \in X.$$

If $f \in \mathcal{H}_K$ then $f(\cdot) = \langle f, \eta(\cdot)\rangle_K$ and a similar calculation leads to the continuity of f. □

The proof of the next result is very close to that in the metric case that appeared in [33] and uses a general version of Arzela–Ascoli's theorem ([23, p. 290]).

Lemma 2.2. *If X is also compact and Hausdorff then the continuity of K implies the compactness of the inclusion map $i : \mathcal{H}_K \hookrightarrow C(X)$. The converse is also true.*

Proof. If K is continuous then it is bounded and there exists $M > 0$ such that

$$|f(x)| = \langle f, K^x \rangle_K \leq \sqrt{K(x,x)}\|f\|_K \leq M\|f\|_K, \quad x \in X, \quad f \in \mathcal{H}_K.$$

It follows that the inclusion map is bounded. Since $\mathcal{H}_K \subset C(X)$ and

$$|f(x) - f(y)|^2 = |\langle f, K^x - K^y \rangle_K|^2 \leq \|f\|_K^2 \|K^x - K^y\|_K^2, \quad x, y \in X, \quad f \in \mathcal{H}_K.$$

Equation (2.1) may be used to see that every bounded set of $\mathcal{H}_K$ is equi-continuous ([23, p. 276]). Now Arzela–Ascoli's theorem guarantees the compacity of the inclusion map (see also Theorem 4.6 in [23, p. 283]). To the converse, if i is compact, B is the unitary closed ball in $\mathcal{H}_K$ and $x, y \in X$, then

$$\sup_{f \in B} |\langle f, K^x - K^y \rangle_K| = \sup_{f \in B} |f(x) - f(y)| = \|K^x - K^y\|_K = \|\eta(x) - \eta(y)\|_K.$$

It follows from Arzela–Ascoli's theorem that B is equi-continuous and hence η is continuous. Lemma 2.1 may be used to ensure that K is also continuous. □

Lemma 2.3 below is essentially proved in [26, p.36]. It reveals that convergence in the topology of $\mathcal{H}_K$ is not too far from uniform convergence on compact subsets of X. The result follows from the inequality

$$|f_n(x) - f(x)| = |\langle f_n - f, K^x \rangle_K| \leq \|f_n - f\|_K \kappa(x)^{1/2}, \quad x \in X,$$

a direct consequence of the reproducing property and the Cauchy–Schwarz inequality.

Lemma 2.3. *If $\{f_n\}$ converges to f in the topology of $\mathcal{H}_K$ then the convergence is actually uniform on any subset A of X in which κ is bounded.*

Next, we detail how one can embed $\mathcal{H}_K$ into $C(X)$ under weaker assumptions. Additional structure on X is needed.

Theorem 2.4. *If X is either a first countable or a locally compact topological space, every function K^x is continuous and the restriction of κ to every compact subset of X is bounded then $\mathcal{H}_K$ is a subset of $C(X)$.*

Proof. We will show that if $\{f_n\}$ converges to f in the topology of $\mathcal{H}_K$ and every f_n is of the form $f_n = \sum_{i=1}^{j_n} c_i K^{x_i}$, $c_i \in \mathbb{C}$, $x_i \in X$, then f is continuous. Let A be a compact subset of X. If κ is bounded in A, Lemma 2.3 implies that f_n is uniformly convergent to f in A. If X is first countable we use Theorem 1.1 in [23] to obtain the continuity of f. If X is locally compact, the same conclusion can be reached with the help of Theorem 10.6 in [23]. □

Clearly, the continuity of K is an ideal replacement for the two assumptions involving K mentioned in the statement of the theorem.

In Learning Theory, the continuity and compactness of the inclusion map $i : \mathcal{H}_K \hookrightarrow C(X)$ are used to prove the existence of the so-called uniform covering numbers ([33]). Such properties of i are always guaranteed when the space X is metric.

Next, we embed $\mathcal{H}_K$ into $L^2(X,\nu)$.

Proposition 2.5. *Assume every function in $\mathcal{H}_K$ is ν-measurable. If κ belongs to $L^1(X,\nu)$ then $\mathcal{H}_K \subset L^2(X,\nu)$. In that case, the inclusion map $i : \mathcal{H}_K \hookrightarrow L^2(X,\nu)$ has norm at most $\|\kappa\|_1^{1/2}$.*

Proof. It suffices to see that

$$\|f\|_2 \leq \|\kappa\|_1^{1/2}\|f\|_K, \quad f \in \mathcal{H}_K.$$

□

Since the measure ν is a Borel (complete) measure, the measurability assumption in the previous proposition can be dropped whenever $\mathcal{H}_K \subset C(X)$. In particular, that is the case when K is continuous. Also, if κ belongs to $L^1(X,\nu)$ then the well known inequality

$$|K(x,y)|^2 \leq \kappa(x)\kappa(y), \quad x,y \in X,$$

implies that K^x belogs to $L^2(X,\nu)$.

Another important issue is to describe conditions under which $\mathcal{H}_K$ contains a copy of the range of $\mathcal{K}$. Proposition 2.6 gives an answer to that.

Proposition 2.6. *Assume every function in $\mathcal{H}_K$ is ν-measurable. If κ belongs to $L^1(X,\nu)$ then the range of $\mathcal{K}$ is a subset of $\mathcal{H}_K$. In that case,*

$$\langle \mathcal{K}(f), g\rangle_K = \langle f, g\rangle_2, \quad g \in \mathcal{H}_K, \quad f \in L^2(X,\nu).$$

Proof. Let $f \in L^2(X,\nu)$ and consider the linear functional $\psi_f : \mathcal{H}_K \to \mathbb{C}$ given by $\psi_f(g) = \langle g, f\rangle_2$, $g \in \mathcal{H}_K$. Applying the Cauchy–Schwarz inequality, we deduce that

$$|\psi_f(g)| \leq \|f\|_2\|\kappa\|_1^{1/2}\|g\|_K, \quad g \in \mathcal{H}_K,$$

that is, ψ_f is bounded. The Riesz representation theorem implies the existence of $h \in \mathcal{H}_K$ such that

$$\psi_f(g) = \langle g, f\rangle_2 = \langle g, h\rangle_K, \quad g \in \mathcal{H}_K.$$

In particular,

$$h(x) = \langle h, K^x\rangle_K = \langle f, K^x\rangle_2 = \mathcal{K}(f)(x), \quad x \in X,$$

and, consequently, $\mathcal{K}(f) \in \mathcal{H}_K$. The equality in the statement of the proposition is implicit in the arguments above. □

Versions of the above results that hold in a metric setting can be found in [31, 33]. We advise that the assumptions used there are stronger than ours. Finally, we observe that in many particular settings, a relevant problem that requires further analysis is that of finding reasonable conditions in order to embed H_K in spaces of smooth functions, whenever the generating kernel K carries a smoothness assumption. Here, the smoothness in $\mathcal{H}_K$ and that of K need to agree somehow. We will discuss some of that ahead.

3. Results along Mercer's theorem

We now change direction and describe results related to series representations for K, $\mathcal{K}$ and the square root $\mathcal{K}^{1/2}$ of $\mathcal{K}$ when K is at least positive definite. We will assume X is a topological space endowed with a strictly positive measure ν, that is, a (complete) Borel measure on X for which two properties hold: every open nonempty subset of X has positive measure and every $x \in X$ belongs to an open subset of X having finite measure. The need for the assumptions above on X and ν arise in technical arguments (see also [15] for more details). Some assumptions on K will be added in each instance. In a certain sense, the results to be presented here can be seen as applications, extensions and generalizations of others proved in [14, 15, 31, 33].

An interesting situation occurs when the kernel K is $L^2(X,\nu)$-*positive definite*, that is, when $\mathcal{K}$ is bounded and

$$\langle \mathcal{K}(f), f\rangle_2 \geq 0, \quad f \in L^2(X,\nu).$$

In that case, if the function $x \in X \to K^x \in L^2(X,\nu)$ is continuous, the formula

$$\mathcal{K}(f)(x) = \langle f, K^x\rangle_2, \quad f \in L^2(X,\nu), \quad x \in X,$$

shows at once that the range of $\mathcal{K}$ is a subset of $C(X)$. If the compactness of $\mathcal{K}$ is assured then a quite general version of Mercer's theorem along the lines of those proved in [13, 15, 31] holds for $\mathcal{K}$. Precisely, one can deduce that $\mathcal{K}$ is self-adjoint and possesses a $L^2(X,\nu)$-convergent spectral representation in the form

$$\mathcal{K}(f) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\langle f, \phi_n\rangle_2 \phi_n, \quad f \in L^2(X,\nu),$$

in which $\{\phi_n\}$ is an orthonormal subset of $L^2(X,\nu)$ and $\{\lambda_n(\mathcal{K})\}$ decreases to 0. Also, since every function in the sequence $\{\lambda_n(\mathcal{K})\phi_n\}$ is continuous then one can show that K has a series representation in the form

$$K(x,y) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\phi_n(x)\overline{\phi_n(y)}, \quad x, y \in X, \tag{3.1}$$

Further, the convergence in both series is absolute and uniform on compact subsets. More details on these facts are to be found in [14, 15].

The comments above suggest the following definition. A continuous kernel K on X will be termed a *Mercer's kernel* when it possesses a series representation of the form (3.1) in which $\{\phi_n\}$ is an $L^2(X,\nu)$-orthonormal sequence of continuous functions on X, $\{\lambda_n(\mathcal{K})\}$ decreases to 0 and the series converges uniformly on compact subsets of $X \times X$. We will assume the representation (3.1) of a Mercer's kernel is such that $\lambda_n(\mathcal{K})\phi_n \neq 0$ for all n and advise the reader that the results to come can be adapted to hold in the case when the series above becomes a finite sum. A Mercer's kernel is necessarily $L^2(X,\nu)$-positive definite. In particular, it is also positive definite in the usual sense. In addition, by using Lemma 2.1, it is promptly seen that the operator $\mathcal{K}$ has all the properties mentioned in the previous paragraph.

The actual verification that a continuous and positive definite kernel is a Mercer's kernel is not an easy task and it does not become easier, even when X is either locally compact or possesses the compactness structure adopted in [31]. An ideal general setting for that to happen is described in the first result in this section.

Theorem 3.1. *Let K be continuous and $L^2(X,\nu)$-positive definite on X. If κ belongs to $L^1(X,\nu)$ then K is a Mercer's kernel.*

Proof. Assume $\kappa \in L^1(X,\nu)$. Since K is positive definite in the usual sense then

$$\int_X |K(x,y)|^2 d(\nu\times\nu)(x,y) \leq \left(\int_X \kappa(x)d\nu(x)\right)^2 < \infty,$$

that is, $K \in L^2(X\times X, \nu\times\nu)$. In particular, $\mathcal{K}$ is compact and self-adjoint. Invoking Lemma 2.1 and Proposition 2.6, we deduce that the range of $\mathcal{K}$ is a subset of $C(X)$. As so, the usual spectral theorem for compact and self-adjoint operators on Hilbert spaces informs us that $\mathcal{K}$ has an $L^2(X,\nu)$-convergent series representation in the form

$$\mathcal{K}(f) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\langle f, \phi_n\rangle_2 \phi_n, \quad f \in L^2(X,\nu), \tag{3.2}$$

where $\{\lambda_n(\mathcal{K})\}$ decreases to 0 and $\{\phi_n\}$ is $L^2(X,\nu)$-orthonormal. In addition, we can assume that each ϕ_n is continuous. The rest of the proof follows well-known arguments which we reproduce for the convenience of the reader. For each $p \geq 1$, the formula

$$K_p(x,y) = K(x,y) - \sum_{n=1}^{p} \lambda_n(\mathcal{K})\phi_n(x)\overline{\phi_n(y)}, \quad x,y \in X,$$

defines a continuous element of $L^2(X\times X, \nu\times\nu)$. Standard computations show that it is positive definite. As so, $K_p(x,x) \geq 0$, $x \in X$, that is,

$$\sum_{n=1}^{p} \lambda_n(\mathcal{K})|\phi_n(x)|^2 \leq K(x,x), \quad x \in X.$$

On the other hand, the inequality

$$\left|\sum_{n=p}^{p+q} \lambda_n(\mathcal{K})\langle f,\phi_n\rangle_2\phi_n(x)\right|^2 \leq \lambda_1(\mathcal{K}) \sup_{y\in Y} K(y,y) \sum_{n=p}^{p+q} |\langle f,\phi_n\rangle_2|^2, \quad x \in Y,$$

holds whenever $Y \subset X$ and $q,p \geq 1$. As so, the convergence of the series in (3.2) is uniform on those subsets of X on which κ is bounded. The inequality

$$\left|\sum_{n=p}^{p+q} \lambda_n(\mathcal{K})\phi_n(x)\overline{\phi_n(z)}\right|^2 \leq K(z,z)\sum_{n=p}^{p+q} \lambda_n(\mathcal{K})|\phi_n(x)|^2, \quad x,z \in X, \quad p,\, q \geq 1,$$

holds due to the Cauchy–Schwarz inequality while the continuity of K guarantees that every $y \in X$ has an open neighborhood $\mathcal{O}_y \subset X$ for which

$$\kappa(z) \leq \kappa(y) + 1, \quad z \in \mathcal{O}_y.$$

Thus, if x is held fixed, the Cauchy criterion for convergence implies the uniform convergence of the series $\sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\phi_n(x)\overline{\phi_n(z)}$ on $\mathcal{O}_y$, to a function $G^x : X \to \mathbb{C}$. That means that $G^x \in C(X)$. Hence,

$$\int_X G^x(y) f(y)\, d\nu(y) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\langle f, \phi_n\rangle_2 \phi_n(x) = \mathcal{K}(f)(x), \quad f \in L^2(X,\nu), \quad x \in X,$$

that is,

$$\int_X [G^x(y) - K(x,y)]\, f(y)\, d\nu(y) = 0, \quad f \in L^2(X,\nu), \quad x \in X.$$

The basic assumptions on X and ν enter in the deduction of

$$\sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\phi_n(x)\overline{\phi_n(y)} = K(x,y), \quad x, y \in X,$$

while a help of Dini's theorem leads to

$$\sum_{n=1}^{\infty} \lambda_n(\mathcal{K})|\phi_n(x)|^2 = K(x,x), \quad x \in X,$$

with uniform and absolute convergence on compact subsets of X. Finally, by applying the Cauchy criterion for uniform convergence and the Cauchy–Schwarz inequality we obtain the uniform and absolute convergence of the series (3.4) on compact subsets of $X \times X$. □

Since Theorem 3.1 does not use either compactness or a metric structure on X, we like to think it is a significant improvement to many other similar results found in the literature (see [3, 12, 14, 15, 18, 22, 24, 31] for example). It also produces other relevant consequences in all the areas surrounding the theory of integral operators generated by positive definite kernels. For instance, a simple application of the monotone convergence theorem reveals that $\|\kappa\|_1 = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})$. Such equality reduces itself to what is called the basic decay rate for the sequence of eigenvalues of $\mathcal{K}$, since it implies $\lambda_n(\mathcal{K}) = o(n^{-1})$ as $n \to \infty$ ([16, 17]). Better decay rates can be reached when we restrict the setting and add smoothness assumptions on K. As a matter of fact this is an interesting area of research that has not found its end yet. Some recent results related to such problem can be found in [3, 8, 14, 15, 17] and many others quoted there. There are plenty of questions still to be analyzed along these lines.

Moving on, we now discuss a method to construct an orthonormal basis for the Hilbert space $\mathcal{H}_K$ when K is a Mercer's kernel.

Theorem 3.2. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. Assume that κ belongs to $L^1(X,\nu)$. With the notation in (3.1), an orthonormal basis to $\mathcal{H}_K$ is the set* $\left\{\sqrt{\lambda_n(\mathcal{K})}\phi_n : n = 1, 2, \ldots\right\}$.

Proof. The operator $\mathcal{K}$ is representable as described in the beginning of the section. As so,

$$\mathcal{K}(\phi_n) = \lambda_n(\mathcal{K})\phi_n, \quad n = 1, 2, \ldots,$$

and, consequently, each ϕ_n belongs to $\mathcal{H}_K \cap L^2(X,\nu)$. If κ belongs to $L^1(X,\nu)$, Proposition 2.6 can be used in the deduction of

$$\langle \lambda_n(\mathcal{K})\phi_n, \phi_m\rangle_K = \langle \mathcal{K}(\phi_n), \phi_m\rangle_K = \langle \phi_n, \phi_m\rangle_2 = \delta_{m,n}.$$

Consequently, $\{\lambda_n(\mathcal{K})^{1/2}\phi_n : n = 1, 2, \ldots\}$ is an orthonormal subset of $\mathcal{H}_K$. To complete the proof, first we use the reproducing property $\phi_n(x) = \langle \phi_n, K^x\rangle_K$, $x \in X$, to see that

$$K^x = K(\cdot, x) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\overline{\phi_n(x)}\,\phi_n = \sum_{n=1}^{\infty} \langle K^x, \phi_n\rangle_K \lambda_n(\mathcal{K})\phi_n, \quad x \in X.$$

Now, if $f \in \mathcal{H}_K$ satisfies $\langle f, \phi_n\rangle_K = 0$, $n = 1, 2, \ldots$ then

$$f(x) = \left\langle f, \sum_{n=1}^{\infty} \langle K^x, \phi_n\rangle_K \lambda_n(\mathcal{K})\phi_n \right\rangle_K = \sum_{n=1}^{\infty} \langle f, \phi_n\rangle_K \lambda_n(\mathcal{K})\phi_n(x) = 0, \quad x \in X.$$

Therefore, $\{\lambda_n(\mathcal{K})^{1/2}\phi_n : n = 1, 2, \ldots\}$ is also a basis of $\mathcal{H}_K$. □

To finalize the section, we will use the previous theorem to characterize $\mathcal{H}_K$ as a special separable subspace of $L^2(X,\nu)$. Before that, we need to list some information regarding the square root of the integral operator generated by a Mercer's kernel.

If K is a Mercer's kernel, basic functional analysis reveals that $\mathcal{K}$ has a unique positive square root $\mathcal{K}^{1/2} : L^2(X,\nu) \to L^2(X,\nu)$ satisfying $\mathcal{K}^{1/2}(\phi_n) = \lambda_n(\mathcal{K})^{1/2}\phi_n$, $n = 1, 2, \ldots$. It can also be shown that

$$\mathcal{K}^{1/2}(f)(x) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^{1/2}\langle f, \phi_n\rangle_2 \phi_n(x), \quad f \in L^2(X,\nu), \quad x \in X,$$

(with uniform convergence on compact subsets of X) and that the range of $\mathcal{K}^{1/2}$ contains continuous functions only (the proof is similar to that in the metric case described in [15]).

For a subset B of $L^2(X,\nu)$, we will write $\overline{B}$ to denote the closure of B in $L^2(X,\nu)$ while $B^\perp$ will denote the orthogonal complement of B in that space. In addition, we will write $[B]$ to denote the linear span of B.

Theorem 3.3. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. If κ belongs to $L^1(X,\nu)$ then $\mathcal{K}^{1/2}$ settles an isometric isomorphism between $\overline{[K^x : x \in X]}$ and $\mathcal{H}_K$.*

Proof. Assume κ belongs to $L^1(X,\nu)$. Since K is a Mercer's kernel, consider the set $\{\phi_n : n = 1, 2, \ldots\}$ from the representation (3.1) once again. Since $[\{K^x : x \in X\}] \subset \overline{[\{\phi_n\}]}$, the inclusion $\overline{[\{K^x : x \in X\}]} \subset \overline{[\{\phi_n\}]}$ follows at once. An immediate consequence is the inclusion $\overline{[\{\phi_n\}]}^\perp \subset \overline{[\{K^x : x \in X\}]}^\perp$. Going the other way around, if $f \in L^2(X,\nu)$ is orthogonal to $\overline{[\{K^x : x \in X\}]}$ then

$$\mathcal{K}(f)(x) = \langle f, K^x\rangle_2 = 0, \quad x \in X,$$

that is, $\mathcal{K}(f) = 0$. Proposition 2.5 implies that

$$\langle f, \phi_n\rangle_2 = \langle \mathcal{K}(f), \phi_n\rangle_K = 0, \quad n = 1, 2, \ldots.$$

Due to the orthonormality of $\{\phi_n\}$ in $L^2(X,\nu)$, we deduce that $\langle f, g\rangle_2 = 0$, whenever $g \in \overline{[\{\phi_n\}]}$. It is now clear that $\overline{[\{\phi_n\}]}^{\perp} \supset \overline{[\{K^x : x \in X\}]}^{\perp}$. To complete the proof, first observe that if $g \in \mathcal{H}_K$, then

$$g = \sum_{n=1}^{\infty} \beta_n \lambda_n(\mathcal{K})^{1/2}\phi_n = \mathcal{K}^{1/2}(f),$$

with $f = \sum_{n=1}^{\infty} \beta_n\phi_n \in \overline{[\{K^x : x \in X\}]}$. Finally, the equality

$$\mathcal{K}^{1/2}(f) = \sum_{n=1}^{\infty} \alpha_n \lambda_n(\mathcal{K})^{1/2}\phi_n, \qquad f = \sum_{n=1}^{\infty} \alpha_n\phi_n \in \overline{[\{K^x : x \in X\}]}$$

and an application of Parseval's identity leads to $\|\mathcal{K}^{1/2}(f)\|_K^2 = \sum_{n=1}^{\infty} |\alpha_n|^2 = \|f\|_2^2$. The proof is complete. □

The following complement of the previous theorem refines Proposition 2.5.

Theorem 3.4. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. If κ belongs to $L^1(X,\nu)$ then the inclusion map $i : \mathcal{H}_K \hookrightarrow L^2(X,\nu)$ has norm at most $\lambda_1(\mathcal{K})^{1/2}$.*

Proof. With the same notation used in Theorem 3.3, let f be of the form

$$f = \sum_{i=1}^{m} \alpha_i K^{x_i} = \sum_{n=1}^{\infty} \left(\lambda_n(\mathcal{K})^{1/2} \sum_{i=1}^{m} \alpha_i \overline{\phi_n(x_i)} \right) \lambda_n(\mathcal{K})^{1/2}\phi_n,$$

in which $\{\phi_n : n = 1, 2, \ldots\}$ comes from Mercer's representation (3.1). If κ belongs to $L^1(X,\nu)$, then $f \in \mathcal{H}_K \cap L^2(X,\nu)$ and Bessel's inequality implies that

$$\|f\|_2^2 = \sum_{n=1}^{\infty} \left| \lambda_n(\mathcal{K})^{1/2} \sum_{i=1}^{m} \alpha_i \overline{\phi_n(x_i)} \right|^2 \lambda_n(\mathcal{K}) \le \lambda_1(\mathcal{K}) \|f\|_K^2.$$

To conclude the proof it suffices to observe that $[K^x : x \in X]$ is both, a subset of $L^2(X,\nu) \cap \mathcal{H}_K$ and dense in $\mathcal{H}_K$. □

The results in this section contain a certain degree of generality we believe is not transposable. But that should deserve a solid reasoning which we don't have at this moment. A possible line of investigation is to move out one of the primary assumptions and to analyze what can be done in order to maintain the outcomes in the previous results. In the next section, we do such an analysis by dropping the integrability of κ.

4. Mercer's theory without integrability of κ

As we have seen, if K is continuous and $L^2(X,\nu)$-positive definite, then the integrability of κ implies the square integrability of K and hence the compactness of $\mathcal{K}$. However, it is plausible that we can have a situation in which $\mathcal{K}$ is not compact, $\mathcal{H}_K$ is still composed of continuous functions and containing the range of $\mathcal{K}$ (see [20] for a such a case). In a certain sense, this situation justifies the pertinency of the analysis in this section.

The basic assumptions on X and ν as in the previous sections persist here. The first result refers to the action of $\mathcal{K}$ on elements of $L^2_c(X,\nu)$, the set of all functions in $L^2(X,\nu)$ having compact support.

Proposition 4.1. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. Then $\mathcal{K}$ takes $L^2_c(X,\nu)$ into $\mathcal{H}_K$ and*

$$\langle \mathcal{K}(f), g\rangle_K = \langle f, g\rangle_2, \quad f \in L^2_c(X,\nu), \quad g \in \mathcal{H}_K.$$

Proof. The proof relies on the very same arguments used in the proof of Proposition 2.5. The estimation on the linear functional ψ takes the form

$$|\psi_f(g)|^2 \leq \|f\|_2^2 \|g\|_K^2 \int_Y k(x)d\nu(x),$$

where Y denotes the support of f which we know is a subset of X of finite measure. The remaining details can be easily accomplished. □

Proposition 4.2 below takes care of the extension of the previous property to $\overline{L^2_c(X,\nu)}$.

Proposition 4.2. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. If a sequence $\{f_n\}$ in $L^2_c(X,\nu)$ converges to f in $L^2(X,\nu)$ then $\mathcal{K}(f)$ belongs to $\mathcal{H}_K \cap C(X)$ and*

$$\langle \mathcal{K}(f), g\rangle_K = \lim_{n\to\infty} \langle f_n, g\rangle_2, \quad g \in \mathcal{H}_K.$$

Proof. Let $\{f_n\} \subset L^2_c(X)$. The previous proposition implies that each $\mathcal{K}(f_n)$ belongs to $\mathcal{H}_K$ and

$$\|\mathcal{K}(f_n) - \mathcal{K}(f_m)\|_K^2 = \langle f_n - f_m, \mathcal{K}(f_n) - \mathcal{K}(f_m)\rangle_2 \leq \|f_n - f_m\|_2 \|\mathcal{K}(f_n) - \mathcal{K}(f_m)\|_2,$$

for all $m, n \geq 1$. If $\{f_n\}$ converges to f in $L^2(X,\nu)$, the continuity of $\mathcal{K}$ implies that $\{\mathcal{K}(f_n)\}$ converges to $\mathcal{K}(f)$ in $L^2(X,\nu)$ while the previous inequality reveals that $\{\mathcal{K}(f_n)\}$ is a Cauchy sequence in $\mathcal{H}_K$. As so, it converges there to the same limit $\mathcal{K}(f)$. Applying Lemma 2.1, we conclude that the convergence is uniform on compact subsets of X. It follows that $\mathcal{K}(f) \in \mathcal{H}_K$ and

$$\langle \mathcal{K}(f), g\rangle_K = \lim_{n\to\infty} \langle \mathcal{K}(f_n), g\rangle_K = \lim_{n\to\infty} \langle f_n, g\rangle_2, \quad g \in \mathcal{H}_K.$$

The proof is complete. □

An alternative version of Theorem 3.1 is as follows.

Theorem 4.3. *Let K be a continuous and $L^2(X,\nu)$-positive definite kernel on X. If $L^2_c(X,\nu)$ is dense in $L^2(X,\nu)$ and $\mathcal{K}$ is compact then K is a Mercer's kernel.*

Proof. It is a consequence of Proposition 4.2 and an adaptation of the procedure used in the proof of Theorem 3.1. □

The additional denseness requirement in the statement of Theorem 4.3 may appear unexpected. What is important to notice is that such property holds automatically in at least two important situations, that in which X is locally compact and ν is a Radon measure and the other in which X is a metric space possessing a compact structure as described in [31]. To exhibit a concrete setting where the denseness assumption does not hold is not easy and will not be done here.

Another quite interesting point to be noticed by the critical reader is that Theorem 4.3 implies that the set of assumptions used in Section 3 in [31] is redundant. Also, it shows that some of the assumptions used in [3, 24] are too demanding. We do not intend to go any further on that in the present work, leaving any additional analysis to the interested reader.

Theorem 4.4. *Let K be a Mercer's kernel on X. Assume that $L_c^2(X,\nu)$ is dense in $L^2(X,\nu)$. If $\{\phi_n : n = 1, 2, \ldots\}$ is the orthonormal set provided by (3.1) then the set $\{\lambda_n(\mathcal{K})^{1/2}\phi_n : n = 1, 2, \ldots\}$ is an orthonormal basis of $\mathcal{H}_K$.*

Proof. Since each ϕ_n is an element of $\mathcal{H}_K \cap L^2(X,\nu)$ we can modify the proof of Theorem 3.3, replacing Proposition 2.6 with Proposition 4.2, to achieve the orthonormality assertion. $\square$

A way to identify $\mathcal{H}_K$ with a subspace of $L^2(X,\nu)$ is described in our next result.

Theorem 4.5. *Let K be a Mercer's kernel on X. Assume that $L_c^2(X,\nu)$ is dense in $L^2(X,\nu)$. Then $\mathcal{K}^{1/2}$ settles an isometric isomorphism between $\overline{[\{K^x : x \in X\}]}$ and $\mathcal{H}_K$. Also, the inclusion map $i : \mathcal{H}_K \hookrightarrow L^2(X,\nu)$ is bounded and its norm is at most $\lambda_1(\mathcal{K})^{1/2}$.*

Proof. It is sufficient to see that

$$\|K^x\|_2^2 = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^2 |\phi_n(x)|^2 \leq \lambda_1(\mathcal{K})\kappa(x) < \infty, \quad x \in X,$$

and to repeat the arguments used in the proofs of theorems 3.3 and 3.4, replacing Proposition 2.6 with Proposition 4.2. $\square$

We close the section with another representation for the reproducing kernel Hilbert space $\mathcal{H}_K$. A particular version of this result in the case when X is a subset of an Euclidean space can be found in [9].

Theorem 4.6. *Let us assume the setting in either Theorem 3.2 or Theorem 4.4. Then the elements in $\mathcal{H}_K$ can be represented in the form*

$$f = \sum_{n=1}^{\infty} \langle f, \phi_n\rangle_2 \, \phi_n, \quad \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^{-1} |\langle f, \phi_n\rangle_2|^2 < \infty. \tag{4.1}$$

The convergence in the first series is absolute and uniform on compact subsets of X. In particular,

$$\langle f, g\rangle_K = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^{-1} \langle f, \phi_n\rangle_2 \langle \phi_n, g\rangle_2, \quad f, g \in \mathcal{H}_K.$$

Proof. Under the conditions stated, if $f \in L^2(X,\nu)$ has the description in (4.1) then it is easily seen that $f \in \mathcal{H}_K$. On the other hand, Theorem 3.2 (respect.

Theorem 4.4) and Proposition 2.6 (respect. 4.2), lead to a representation for each $f \in \mathcal{H}_K$ in the form

$$f = \sum_{n=1}^{\infty} \langle f, \lambda_n(\mathcal{K})^{1/2}\phi_n\rangle_K \lambda_n(\mathcal{K})^{1/2}\phi_n = \sum_{n=1}^{\infty} \langle f, \mathcal{K}(\phi_n)\rangle_K \phi_n = \sum_{n=1}^{\infty} \langle f, \phi_n\rangle_2 \, \phi_n.$$

The inequality

$$\left|\sum_{n=j}^{\infty} \langle f, \lambda_n(\mathcal{K})^{1/2}\phi_n\rangle_K \lambda_n(\mathcal{K})^{1/2}\phi_n(x)\right|^2 \leq \|f\|_K^2 \kappa(x), \quad x \in X, \quad j = 1, 2, \ldots,$$

produces the absolute and uniformly convergence on compact subsets of X. The continuity of the inner product of $\mathcal{H}_K$ plus the previous arguments lead to the final statement in the theorem. □

5. RKHS of Lipchitzian functions

In this section we intend to describe conditions under which $\mathcal{H}_K$ and the range of both $\mathcal{K}$ and $\mathcal{K}^{1/2}$ are spaces of lipschitzian functions. For obvious reasons, we need to assume that X is metrizable while ν continues to be a strictly positive measure on X. The metric in X will be written as d.

For $\alpha > 0$, the symbol $Lip^\alpha(X, \nu)$ will designate the class of all kernels K on X satisfying the following condition: there exists $\delta > 0$ and a function $A : X \to [0, \infty]$ in $L^2(X, \nu)$ such that

$$|K(x, y) - K(z, y)| \leq A(y) d(x, z)^\alpha,$$

whenever $x, y, z \in X$ and $d(x, z) < \delta$.

Proposition 5.1 below reveals a basic property the bounded operator $\mathcal{K}$ has when K comes from $Lip^\alpha(X, \nu)$. The reader should notice that the Lipschitz condition just introduced does not guarantee continuity of K.

Proposition 5.1. *Let K be a kernel in $L^2(X \times X, \nu \times \nu)$. If K belongs to $Lip^\alpha(X, \nu)$ then the range of $\mathcal{K}$ is entirely composed of usual α-Lipschitzian functions.*

Proof. If $K \in Lip^\alpha(X, \nu) \cap L^2(X \times X, \nu \times \nu)$ and $f \in L^2(X, \nu)$ then it is easily seen that

$$\begin{aligned} |\mathcal{K}(f)(x) - \mathcal{K}(f)(z)| &\leq \int_X |K(x, y) - K(z, y)||f(y)| d\nu(y) \\ &\leq d(x, z)^\alpha \int_X A(y)|f(y)| d\nu(y), \end{aligned}$$

whenever $x, z \in X$ and $d(x, z) < \delta$, in which A is the function that realizes the definition above for K. An application of Hölder's inequality leads to

$$|\mathcal{K}(f)(x) - \mathcal{K}(f)(z)| \leq \|A\|_2 \|f\|_2 d(x, z)^\alpha,$$

whenever $x, z \in X$ and $d(x, z) < \delta$. This is precisely what the usual definition of an α-Lipschitzian function requires. □

The following concept is a variation of the previous one. It seems to be more suitable in the analysis of decay rates for the eigenvalues of $\mathcal{K}$ in some special settings (see [14, 15, 17] for example). For $\alpha > 0$, the kernel K is said to be *α-Lipschtizian in the diagonal* of $X \times X$ when there exists a positive constant A so that

$$|K(x,x) - K(x,y)| \leq A\, d(x,y)^{\alpha},$$

whenever $x, y \in X$ and $d(x,y) < \delta$.

If a kernel K is α-Lipschtizian in the diagonal of $X \times X$ then κ is continuous and K^x is continuous at $y = x$. In addition the following result holds.

Proposition 5.2. *Let K be positive definite on X. If it is α-Lipschtizian in the diagonal of $X \times X$ then $\mathcal{H}_K$ is entirely composed of $\alpha/2$-Lipschitzian functions.*

Proof. Assume K is α-Lipschitzian in the diagonal of $X \times X$. Invoking the reproducing propositionerty in $\mathcal{H}_K$ and (2.1), we immediately obtain

$$\begin{aligned} |f(x) - f(y)| &\leq \|f\|_K(|K(x,x) - K(x,y)| + |K(y,y) - K(y,x)|)^{1/2} \\ &\leq \|f\|_K (2A)^{1/2} d(x,y)^{\alpha/2}, \quad f \in \mathcal{H}_K, \end{aligned}$$

whenever $x, y \in X$ and $d(x,z) < \delta$. The result follows. □

We now recall that for a continuous and $L^2(X,\nu)$-positive definite kernel K, the operator $\mathcal{K}$ possesses a unique positive square root $\mathcal{K}^{1/2}$. Depending on the setting, the following additional propositionerties may also be true: $\mathcal{K}^{1/2}$ is a well defined operator on $L^2(X,\nu)$, it is an integral operator generated by a kernel $K_{1/2}$ on X and the original kernel K can be recovered from $K_{1/2}$ (see [15]), that is,

$$\int_X K_{1/2}(x,u)\overline{K_{1/2}(x,v)}\, d\nu(x) = K(v,u), \quad u, v \in X.$$

In that case, we have the following additional result.

Proposition 5.3. *If, in addition to what was mentioned in the previous paragraph, K is α-Lipschtizian in the diagonal of $X \times X$, then the range of $\mathcal{K}^{1/2}$ is entirely composed of $\alpha/2$-Lipschitzian functions.*

Proof. If $f \in L^2(X,\nu)$ and K is α-Lipschtizian in the diagonal of $X \times X$, an application of Hölder's inequality leads to

$$\begin{aligned} \left|\mathcal{K}^{1/2}(f)(x) - \mathcal{K}^{1/2}(f)(z)\right| &\leq \int_X |K_{1/2}(x,y) - K_{1/2}(z,y)||f(y)|d\nu(y) \\ &\leq \left(\int_X |K_{1/2}(x,y) - K_{1/2}(z,y)|^2 d\nu(y)\right)^{1/2} \|f\|_2, \end{aligned}$$

for all $x, z \in X$. Using the recovery formula, the integral I appearing above can be estimated in the following way

$$\begin{aligned} I &= \int_X (K_{1/2}(x,y) - K_{1/2}(z,y))(\overline{K_{1/2}(x,y)} - \overline{K_{1/2}(z,y)})d\nu(y) \\ &= \kappa(x) - K(x,z) - K(z,x) + \kappa(z) \\ &\leq 2A\, d(x,z)^{\alpha}, \end{aligned}$$

whenever $x, z \in X$ and $d(x,z) < \delta$. The result follows. □

Preliminary versions of the results described in this section appeared in [13] for some specific choices of X. An interesting problem which we think has no answer so far is to determine the most general setting under which $\mathcal{K}$ possesses a unique square root having some or all the properties mentioned before Proposition 5.3. It seems to be mandatory that the setting include the continuity of K but that requires a clear proof.

6. RKHS of differentiable functions

In this section, we will assume X is an open subset of the Euclidian space $\mathbb{R}^d$ and ν is the usual Lebesgue measure. We will denote by $C^s(X)$ the set of all functions defined on X and possessing continuous partial derivatives up to order s. We will also put $C^\infty(X) = \cap_{s=1}^\infty C^s(X)$. Our concern here is to establish a setting in order to obtain RKHSs entirely composed of differentiable functions on X and to deduce reproducing formulas for the derivatives of functions in the RKHS. While the results described here are based on those in [13], there are many other related research works where similar results are obtained for different choices of X and alternative concepts of differentiability. We mention [32, 34] for results in a setting similar to the one adopted here and [8, 20, 21] for results on the unit sphere through two different notions of differentiability. There are plenty of work to be done along this line of research, the application of the results in other areas being a relevant issue. Proofs and references for the results cited here can be found in [13].

If α is a multi-index in $\mathbb{Z}_+^d$ and f is a locally integrable function in X then the α^{th}-weak derivative of f is a function $D_w^\alpha f$ with domain X satisfying

$$\int_X D_w^\alpha f(x)g(x)d\nu(x) = (-1)^{|\alpha|}\int_X f(x)D^\alpha g(x)d\nu(x), \quad g \in C_0^\infty(X),$$

in which

$$C_0^\infty(X) = \{g \in C^\infty(X) : \mathrm{supp}\,(g) \subset X\},$$

and

$$D^\alpha g := \frac{\partial^{|\alpha|}g}{\partial^\alpha x} = \frac{\partial^{|\alpha|}g}{\partial^{\alpha_1}x_1\partial^{\alpha_2}x_2\ldots\partial^{\alpha_d}x_d}.$$

If $D^\alpha f$ exists and is locally integrable then $D_w^\alpha f$ exists and both derivatives coincide. If s is a nonnegative integer, let us write

$$H^s := \{f \in L^2(X,\nu) :\ D_w^\alpha f \in L^2(X,\nu), \quad |\alpha| \leq s\}.$$

The set H^s becomes a Hilbert space when we endow it with the inner product

$$\langle f, g\rangle_{2,s} := \sum_{0\leq|\alpha|\leq s}\langle D_w^\alpha f, D_w^\alpha g\rangle_2, \quad f,\, g \in H^s,$$

and the set $\{f \in C^s(X) :\ D^\alpha f \in L^2(X,\nu), |\alpha| \leq s\}$ becomes a dense subset of H^s.

If K is a continuous kernel on X defining an element in $L^2(X \times X)$, standard analysis can be used to show that every function in the range of $\mathcal{K}$ has weak derivatives up to order s whenever $s > 0$ and the usual α^{th}-partial derivative $D_x^\alpha K$ of K with respect to the variable x belongs to $L^2(X \times X, \nu \times \nu)$, whenever

$0 \leq |\alpha| \leq s$. In addition, the range of $\mathcal{K}$ is a subset of H^s and the corresponding inclusion of this range into H^s has norm at most

$$\left(\sum_{0\leq |\alpha| \leq s} \|D_x^\alpha K\|_2^2 \right)^{1/2}.$$

If $D_x^\alpha K$ exists and is continuous on X, an application of Leibnitz's rule ([23, p.324]) leads to the formula

$$D^\alpha \mathcal{K}(f)(x) = \int_X D_x^\alpha K(x,y) f(y) d\nu(y), \quad f \in C_0^\infty(X), \quad x \in X.$$

From now on we will assume K is as described in the previous paragraph and will add other assumptions at specific points where they are needed. For multi-indices α and β in $\mathbb{Z}_+^d$ the symbol $D_{xy}^{\alpha\beta} K$ will stand for the β^{th}-partial derivative of $D_x^\alpha K$ with respect to the variable y.

Proposition 6.1. *Assume K is $L^2(X,\nu)$-positive definite on X. If α is a multi-index for which $D_x^\alpha K$ exists and is locally integrable and $D_{xy}^{\alpha\alpha} K$ belongs to $L^2(X \times X, \nu\times\nu)$, then $D_{xy}^{\alpha\alpha} K$ is $L^2(X,\nu)$-positive definite.*

Proof. It suffices to obtain the $L^2(X,\nu)$-positive definiteness for functions in $C_0^\infty(X)$. If f is such one function and $f \otimes \overline{f}(x,y) := f(x)\overline{f(y)}$, $x, y \in X$, Fubini's theorem and the $L^2(X,\nu)$-positive definiteness of K, is all that is needed in order to deduce that

$$\begin{aligned} \int_{X\times X} D_{xy}^{\alpha\alpha} K f \otimes \overline{f} d\nu \times \nu &= (-1)^{|\alpha|} \int_X \left(\int_X D_x^\alpha K(\cdot,y) D^\alpha f(y) d\nu(y) \right) f d\nu \\ &= \int_X \left(\int_X K(x,y) D^\alpha f(y) d\nu(y) \right) D^\alpha f(x) d\nu(x) \end{aligned}$$

is nonnegative. □

A bonus derived from the previous result is as follows ([13]). Its proof can be reached by adapting the methods used in [5, 6, 7].

Proposition 6.2. *Assume K is $L^2(X,\nu)$-positive definite on X. If K belongs to $C^{2s}(X \times X)$ then*

$$\left| D_{xy}^{\alpha\beta} K(x,y) \right|^2 \leq D_{xy}^{\alpha\alpha} K(x,x)\, D_{xy}^{\beta\beta} K(y,y), \quad |\alpha|, |\beta| \leq s, \quad x, y \in X.$$

Also, if κ belongs to $L^1(X,\nu)$ then

$$D_{xy}^{\alpha\beta} K(x,y) = \sum_{n=1}^\infty \lambda_n(\mathcal{K}) D^\alpha \phi_n(x) \overline{D^\beta \phi_n(y)}, \quad x, y \in X, \quad |\alpha|, |\beta| \leq s,$$

with absolute and uniformly convergence on compact subsets of $X \times X$. The ϕ_n comes from the Mercer's representation implied by Theorem 3.1. The range of $\mathcal{K}$ is a subset of $C^s(X)$ and

$$D^\alpha \mathcal{K}(f)(x) = \int_X D_x^\alpha K(x,y) f(y) d\,\nu(y), \quad x \in X, \quad f \in L^2(X,\nu), \quad |\alpha| \leq s.$$

In the statement and proof of the next result we make use of the partial derivative $D_y^\alpha K$ of K with respect to the second variable y.

Theorem 6.3. *Let K be a $L^2(X,\nu)$-positive definite kernel in $C^{2s}(X\times X)$. If κ belongs to $L^1(X,\nu)$ and $|\alpha|\leq s$ then the following assertions hold:*
(i) $\mathcal{H}_K\subset C^s(X)$;
(ii) If $f\in\mathcal{H}_K$ then

$$|D^\alpha f(x)|\leq \left(D_{xy}^{\alpha\alpha}K(x,x)\right)^{1/2}\|f\|_K,\quad x\in X;$$

(iii) The norm of the inclusion map $i:\mathcal{H}_K\to H^s$ is upper bounded by

$$\left(\sum_{0\leq|\alpha|\leq s}\int_X D_{xy}^{\alpha\alpha}K(x,x)d\nu(x)\right)^{1/2};$$

(iv) Each derivative $D_y^\alpha K(\cdot,x)$ belongs to $\mathcal{H}_K$ and

$$D^\alpha f(x)=\left\langle f, D_y^\alpha K(\cdot,x)\right\rangle_K,\quad x\in X,\quad f\in\mathcal{H}_K;$$

(v) If $\{f_n\}$ is a bounded sequence in $\mathcal{H}_K$ and Y is a compact subset of X, then there exists a subsequence $\{f_{n_j}\}$ and $f\in\mathcal{H}_K$ such that $\{D^\alpha f_{n_j}\}$ converges uniformly to $D^\alpha f$ on Y.

Proof. Assume κ belongs to $L^1(X,\nu)$. If $r\geq 1/2$ and $|\alpha|\leq s$, we can employ the Cauchy–Schwarz inequality, Theorem 6.2 and Bessel's inequality to deduce that

$$\begin{aligned}\left|\sum_{n=j}^\infty \lambda_n(\mathcal{K})^r\langle f,\phi_n\rangle_2 D^\alpha\phi_n(x)\right|^2 &\leq \sum_{n=j}^\infty|\langle f,\phi_n\rangle_2|^2\sum_{n=j}^\infty\lambda_n(\mathcal{K})^{2r}|D^\alpha\phi_n(x)|^2\\ &\leq \lambda_1(\mathcal{K})^{2r-1}D_{xy}^{\alpha\alpha}K(x,x)\|f\|_2^2,\end{aligned}$$

for all $x\in X,\quad f\in L^2(X,\nu)$, whenever $j\geq 1$. To handle (i), let $f\in\mathcal{H}_k$. Using the representation for $\mathcal{H}_k$ described in theorems 3.2 and 3.3, we can write

$$f(x)=\mathcal{K}^{1/2}(g)(x)=\sum_{n=1}^\infty\lambda_n(\mathcal{K})^{1/2}\langle g,\phi_n\rangle_2\phi_n(x),\quad x\in X,$$

for some $g\in L^2(X,\nu)$ with $\|f\|_K=\|g\|_2$. In particular, $f\in C^s(X)$. Since

$$|D^\alpha f(x)|^2=\left|\sum_{n=1}^\infty\lambda_n(\mathcal{K})^{1/2}\langle g,\phi_n\rangle_2 D^\alpha\phi_n(x)\right|^2\leq D_{xy}^{\alpha\alpha}K(x,x)\|g\|_2^2,\quad x\in X,$$

assertions (ii) and (iii) follow. Since the formula

$$D_y^\alpha K(\cdot,x)=\sum_{n=1}^\infty\lambda_n(\mathcal{K})^{1/2}\overline{D^\alpha\phi_n(x)}\lambda_n(\mathcal{K})^{1/2}\phi_n,\quad x\in X,$$

holds (see Theorem 6.2), it follows from Theorem 3.2 that $D_y^\alpha K(\cdot,x)\in\mathcal{H}_K$, $x\in X$, and

$$\langle D_y^\alpha K(\cdot,x),D_y^\alpha K(\cdot,y)\rangle_K=D_{xy}^{\alpha\alpha}K(y,x),\quad x,y\in X.$$

Hence, if $f=\mathcal{K}^{1/2}(g)$ for some $g\in L^2(X,\nu)$, then

$$D^\alpha\mathcal{K}^{1/2}(g)(x)=\sum_{n=1}^\infty \lambda_n(\mathcal{K})^{1/2}\langle g,\phi_n\rangle_2 D^\alpha\phi_n(x)=\left\langle \mathcal{K}^{1/2}(g),D_y^\alpha K(\cdot,x)\right\rangle_K,\quad x\in X,$$

and the reproducing property in (iv) follows. In order to take care of (v), let $\{f_n\}$ be bounded in $\mathcal{H}_K$. Due to (iv), we have that

$$\begin{aligned}|D^\alpha f_n(x)-D^\alpha f_n(y)|^2 &= \left|\left\langle f_n, D_y^\alpha K(\cdot,x)-D_y^\alpha K(\cdot,y)\right\rangle_K\right|^2\\ &\leq \left\|D_y^\alpha K(\cdot,x)-D_y^\alpha K(\cdot,y)\right\|_K^2\sup_n\|f_n\|_K^2,\\ &= \left[D_{xy}^{\alpha\alpha}K(x,x)+D_{xy}^{\alpha\alpha}K(y,y)-D_{xy}^{\alpha\alpha}K(x,y)\right.\\ &\qquad\left.-D_{xy}^{\alpha\alpha}K(y,x)\right]\sup_j\|f_{n_j}\|_K^2,\end{aligned}$$

whenever $x,y\in\overline{Y}$ and Y is a bounded subset of X. Thus, recalling that $D_{xy}^{\alpha\alpha}K$ is uniformly continuous on compact subsets of $X\times X$, an application of Ascoli–Arzela's theorem reveals that the restriction of $D^\alpha f_{n_j}$ to $\overline{Y}$ converges uniformly to $D^\alpha f_Y$, for some $f_Y\in C_B^s(\overline{Y})$. To see that f_Y is the restriction to $\overline{Y}$ of a function in $\mathcal{H}_K$, we use the weak compactness of closed balls in $\mathcal{H}_K$ ([25, p.202]), to find a subsequence $\{f_{n_{jl}}\}$ of $\{f_{n_j}\}$, weakly convergent to some f in $\mathcal{H}_K$. Since

$$D^\alpha f_{n_{ji}}(x)-D^\alpha f_{n_{jl}}(x)=\left\langle f_{n_{ji}}-f_{n_{jl}},D_y^\alpha K(\cdot,x)\right\rangle_K,\quad x\in X,\quad l\geq i,$$

we can see that $D^\alpha f_{n_{jl}}$, converges pointwise to $D^\alpha f$. It follows that f_Y is the restriction of f to $\overline{Y}$. □

An easy consequence of this theorem is the next technical result ([13]).

Corollary 6.4. *Let K be a positive definite kernel in $C^{2s}(X\times X)$ for which κ belongs to $L^1(X,\nu)$. If $\{f_n\}$ is a bounded sequence in $\mathcal{H}_K$ then there exists a subsequence $\{f_{n_j}\}$ of $\{f_n\}$ and $f\in\mathcal{H}_K$ such that $\{D^\alpha f_{n_j}\}$ converges uniformly to $D^\alpha f$ on any compact subset of X, whenever $|\alpha|\leq s$.*

We can adapt the results presented in the previous sections to hold for functions defined in the closure $\overline{X}$ of X. We will write $C^s(\overline{X})$ to denote the linear space of all functions in $C^s(X)$ for which all partial derivatives up to order s have a continuous extension to $\overline{X}$. We will write $C_B^s(\overline{X})$ to denote the Banach space of all functions in $C^s(\overline{X})$ which possess bounded partial derivatives, endowed with the norm

$$\|g\|=\max_{0\leq|\alpha|\leq s}\sup_{x\in X}|D^\alpha g(x)|,\quad g\in C_B^s(\overline{X}).$$

Also, we will denote by K_1 the extension of K to $\overline{X}\times\overline{X}$.

Theorem 6.5. *Let K be a positive definite kernel in $C^{2s}(X\times X)$ for which κ belongs to $L^1(X,\nu)$. Assume all the partial derivatives of K are continuously extendable to $\overline{X}\times\overline{X}$ and that each mapping*

$$x\in X\to D_{xy}^{\alpha\alpha}K(x,x),\quad |\alpha|\leq s,$$

is bounded. The following assertions hold:
(i) Both inclusion maps $i : \mathcal{H}_K \to C_B^s(X)$ *and* $i_1 : \mathcal{H}_{K_1} \to C_B^s(\overline{X})$ *have norm at most*

$$\max_{0\leq|\alpha|\leq s} \sup_{x\in X} \left(D_{xy}^{\alpha\alpha} K(x,x)\right)^{1/2};$$

(ii) The image of a bounded and closed subset of $\mathcal{H}_{K_1}$ *by* i_1 *is closed in* $C_B^s(\overline{X})$*. In particular, if either* X *is bounded or* $\lim_{\substack{x\in X\\|x|\to\infty}} k(x) = 0$*, the inclusion map* i_1 *is compact.*

Proof. We can see that $D^\alpha g$ has an extension to $\overline{X}$, for all $g \in \mathcal{H}_K$. Theorem 6.3 implies that

$$|D^\alpha g(x)| \leq \sup_{y\in X} \left(D_{xy}^{\alpha\alpha} K(y,y)\right)^{1/2} \|g\|_K, \quad x \in \overline{X}, \quad g \in \mathcal{H}_{K_1}.$$

It follows that the functions in $\mathcal{H}_{K_1}$ can be seen as functions in $C_B^s(\overline{X})$. As so, the assertions in (i) follow. To finish the proof, let B be a bounded and closed subset of $\mathcal{H}_{K_1}$. If $\{f_n\}$ is a uniformly convergent sequence in B, Theorem 6.3-(v) and Corollary 6.4 imply that its limit f belongs to $\mathcal{H}_{K_1}$. Since B is also weakly compact, there exists a subsequence of $\{f_n\}$ weakly (and pointwise) convergent in B. It follows that f belongs to B, that is, B is a closed set in $C(X)$. If either X is bounded or $\lim_{\substack{x\in X\\|x|\to\infty}} k(x) = 0$ then Corollary 6.4 implies the compactness of B (see also [19, p.132]). □

7. Fourier transform and positive definiteness

In this section we change the focus a little bit to present some results involving both concepts, positive definiteness and the Fourier transform in $\mathbb{R}^m$. Some of the results were motivated by others proved in [4].

The notation and basic properties of the Fourier transform are those from [19]. Here, the usual inner product of two points x, y in the euclidian space $\mathbb{R}^m$ will be written as $x\cdot y$. As so, the *Fourier transform* is the linear mapping $f \in L^1(\mathbb{R}^m) \mapsto \hat{f}$ given by the formula

$$\hat{f}(v) = \int_{\mathbb{R}^m} f(x) e^{-2\pi i x\cdot v}\, dx, \quad v \in \mathbb{R}^m.$$

Since

$$|\hat{f}(v)| \leq \int_{\mathbb{R}^m} |f(x) e^{-2\pi i x\cdot v}|\, dx = \int_{\mathbb{R}^m} |f(x)| dx, \quad v \in \mathbb{R}^m,$$

it is easily seen that the range of the Fourier transform is composed of bounded functions.

Below, we quote an important result from Fourier transform theory ([19, p.253]) due to Plancherel.

Proposition 7.1. *If* $f \in L^1(\mathbb{R}^m) \cap L^2(\mathbb{R}^m)$ *then* $\hat{f} \in L^2(\mathbb{R}^m)$*. Also, there is a unique isometric operator* $\mathcal{F} : L^2(\mathbb{R}^m) \to L^2(\mathbb{R}^m)$ *such that*

$$\mathcal{F}(f) = \hat{f}, \quad f \in L^1(\mathbb{R}^m) \cap L^2(\mathbb{R}^m).$$

The above result shows that the Fourier transform acting on $L^1(\mathbb{R}^m) \cap L^2(\mathbb{R}^m)$ has a unique extension to a linear isometric mapping $L^2(\mathbb{R}^m) \to L^2(\mathbb{R}^m)$. This isometry is actually a unitary map and, consequently, it is possible to speak of Fourier transforms of functions in $L^2(\mathbb{R}^m)$. Thus keeping the notation above for the Fourier transform of functions in $L^2(\mathbb{R}^m)$, the following consequence holds:

$$\langle f, g\rangle_2 = \int_{\mathbb{R}^m} f(x)\overline{g(x)}\, dx = \int_{\mathbb{R}^m} \hat{f}(x)\overline{\hat{g}(x)}\, dx = \langle \hat{f}, \hat{g}\rangle_2, \quad f, g \in L^2(\mathbb{R}^m).$$

Next, we continue with some technical results to be used ahead. We will denote by C_0 the set of all uniformly continuous functions $f : \mathbb{R}^m \to \mathbb{C}$ for which $\lim_{|v|\to\infty} f(v) = 0$. Lemma 7.2 below is proved in [19, p. 249].

Lemma 7.2. *If f belongs to $L^1(\mathbb{R}^m)$ then $\hat{f}$ is an element of C_0.*

Lemma 7.3. *If f belongs to $L^2(\mathbb{R}^m)$ then $\widehat{\overline{f(v)}} = \overline{\hat{f}(-v)}$, $v \in \mathbb{R}^m$.*

Using the usual identification between $\mathbb{R}^m \times \mathbb{R}^m$ and $\mathbb{R}^{2m}$, the Fourier transform of a function f in $L^1(\mathbb{R}^m \times \mathbb{R}^m)$ can be given, via Fubini's theorem, by the formula

$$\hat{f}(v, w) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f(x, y) e^{-2\pi i (x,y)\cdot(v,w)}\, dx\, dy = \int_{\mathbb{R}^{2m}} g(z) e^{-2\pi i z\cdot u}\, dz = \hat{g}(z),$$

in which $z = (x, y)$ and $g(z) = f(x, y)$. With this information in mind, we have the following technical result.

Lemma 7.4. *If ϕ and ψ belongs to $L^2(\mathbb{R}^m)$ and $\phi \otimes \psi(v, w) := \phi(v)\psi(w)$, $v, w \in \mathbb{R}^m$, then*

$$\widehat{\phi \otimes \psi}(v, w) = \hat{\phi}(v)\hat{\psi}(w), \quad v, w \in \mathbb{R}^m.$$

Proof. Let ϕ and ψ be in $L^2(\mathbb{R}^m)$. There are two sequences $\{\phi_n\}$ and $\{\psi_n\}$ in $L^1(\mathbb{R}^m) \cap L^2(\mathbb{R}^m)$ convergent (in $L^2(\mathbb{R}^m)$) to ϕ and ψ, respectively. It is easily seen that the sequence $\{\phi_n \otimes \psi_n\}$ converges to $\phi \otimes \psi$ in $L^2(\mathbb{R}^m \times \mathbb{R}^m)$. Hence, $\phi_n \otimes \psi_n \in L^1(\mathbb{R}^m \times \mathbb{R}^m) \cap L^2(\mathbb{R}^m \times \mathbb{R}^m)$, $n = 1, 2, \ldots$, and

$$\widehat{\phi_n \otimes \psi_n}(v, w) = \hat{\phi}_n(v)\hat{\psi}_n(w), \quad v, w \in \mathbb{R}^m, \quad n = 1, 2, \ldots.$$

The continuity of the Fourier transform implies that $\widehat{\phi \otimes \psi}(v, w) = \hat{\phi}(v)\hat{\psi}(w)$, $v, w \in \mathbb{R}^m$ a.e.. By changing $\widehat{\phi \otimes \psi}$ in a null, if needed, the result follows. $\square$

If the reader is asking himself about what connections may exist between positive definiteness and Fourier transforms, the classical result below may be suggestive (see [30] and many other references).

Theorem 7.5. *A kernel K of the form $K(x, y) = K'(x-y)$, $x, y \in \mathbb{R}^m$, for some function $K' : \mathbb{R}^m \to \mathbb{R}$, is positive definite if and only if there exists a finite Borel measure σ for which*

$$K'(x) = \int_{\mathbb{R}^m} e^{iy\cdot x}\, d\sigma(y), \quad x \in \mathbb{R}^m.$$

To proceed, we introduce the notation

$$\tilde{K}(u,v) := \hat{K}(u,-v), \quad u,v \in \mathbb{R}^m,$$

whenever K belongs to $L^2(\mathbb{R}^m \times \mathbb{R}^m)$. Hence, $\tilde{\mathcal{K}}$ denotes the integral operator generated by $\tilde{K}$.

Theorem 7.6. *If K is a positive definite kernel in $L^2(\mathbb{R}^m \times \mathbb{R}^m)$ then it holds*

$$\widehat{\mathcal{K}(f)} = \tilde{\mathcal{K}}(\hat{f}), \quad f \in L^2(\mathbb{R}^m).$$

Proof. If $K \in L^2(\mathbb{R}^m \times \mathbb{R}^m)$ is positive definite then $\mathcal{K}$ is a compact self-adjoint operator. Using its spectral decomposition

$$\mathcal{K}(f) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K}) \langle f, \phi_n \rangle_2 \phi_n, \quad f \in L^2(\mathbb{R}^m),$$

we can deduce that

$$\widehat{\mathcal{K}(f)} = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K}) \langle \hat{f}, \hat{\phi}_n \rangle_2 \hat{\phi}_n, \quad f \in L^2(\mathbb{R}^m).$$

Since

$$K = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K}) \phi_n \otimes \overline{\phi_n},$$

then

$$\hat{K}(x,y) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K}) \hat{\phi}_n(x) \overline{\hat{\phi}_n(-y)}, \quad x,y \in X \quad \text{a.e},$$

and we can use lemmas 7.3 and 7.4 to conclude that $\widehat{\mathcal{K}(f)} = \tilde{\mathcal{K}}(\hat{f})$, $f \in L^2(\mathbb{R}^m)$. □

An immediate consequence of the previous theorem is a series of results describing clear methods to construct $L^2(\mathbb{R}^m, \nu)$-positive definite kernels from a given one.

Corollary 7.7. *If K belongs to $L^2(\mathbb{R}^m \times \mathbb{R}^m)$ then K is $L^2(\mathbb{R}^m, \nu)$-positive definite if and only if $\tilde{K}$ is so.*

Corollary 7.8. *If K belongs to $L^2(\mathbb{R}^m \times \mathbb{R}^m)$ then K is $L^2(\mathbb{R}^m, \nu)$-positive definite if and only if there exists a $L^2(\mathbb{R}^m, \nu)$-positive definite kernel G such that $K = \tilde{G}$.*

Corollary 7.9. *Let K be an element of $L^2(\mathbb{R}^m \times \mathbb{R}^m)$ and define $K_{(1)} = K$, $K_{(n)} = \tilde{K}_{(n-1)}$, $n = 2, 3, \ldots$. If K is $L^2(\mathbb{R}^m, \nu)$-positive definite then the kernels $K_{(n)}$ are so.*

In the next theorem we register a Mercer's representation for the kernel $\tilde{K}$ when K is a Mercer's kernel on $\mathbb{R}^m$.

Theorem 7.10. *If K is a Mercer's kernel on $\mathbb{R}^m$ and $\kappa^{1/2}$ belongs to $L^1(\mathbb{R}^m)$ then $\tilde{K}$ is a Mercer's kernel with the following series representation derived from that of K:*

$$\tilde{K}(x,y) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})\hat{\phi}_n(x)\overline{\hat{\phi}_n(y)}, \quad x,y \in \mathbb{R}^m.$$

The convergence is absolute and uniform on compact subsets of $\mathbb{R}^m$. The range of $\tilde{\mathcal{K}}$ is a subset of $C_0(\mathbb{R}^m) \cap L^2(\mathbb{R}^m)$.

Proof. If $\kappa^{1/2} \in L^1(\mathbb{R}^m)$ then $K \in L^1(\mathbb{R}^m \times \mathbb{R}^m)$ and

$$|\mathcal{K}(f)(x)| \leq \lambda_1(\mathcal{K})\kappa(x)^{1/2}\|f\|_2, \quad f \in L^2(\mathbb{R}^m), \quad x \in \mathbb{R}^m.$$

As so, the range of $\mathcal{K}$ is a subset of $C(\mathbb{R}^m) \cap L^1(\mathbb{R}^m)$. To close the proof, it suffices to apply Lemma 7.2 and use the series representation for $\tilde{K}$ obtained from that of K, as in the proof of Theorem 7.6. □

Returning to RKHSs, the following result is quite interesting.

Theorem 7.11. *If K is a Mercer's kernel and $\kappa^{1/2}$ belongs to $L^1(\mathbb{R}^m)$ then*

$$\langle \mathcal{K}(f), g\rangle_K = \langle f, g\rangle_2 = \langle \hat{f}, \hat{g}\rangle_2 = \langle \tilde{\mathcal{K}}(\hat{f}), \hat{g}\rangle_{\tilde{K}}, \quad f \in L^2(X,\nu), \quad g \in \mathcal{H}_K.$$

Also,

$$\mathcal{H}_{\tilde{K}} = \{\hat{f} : f \in \mathcal{H}_K\}.$$

Proof. If the assumptions in the statement of the theorem hold, Proposition 4.2 implies that

$$\langle \mathcal{K}(f), g\rangle_K = \langle f, g\rangle_2, \quad f \in L^2(\mathbb{R}^m), \quad g \in \mathcal{H}_K.$$

Hence, the first assertion of the theorem follows from theorems 7.1 and 7.10. As for the last one, let f be in $\mathcal{H}_K$. Since K is a Mercer's kernel, there exists $g \in L^2(\mathbb{R}^m)$ such that

$$f(x) = \mathcal{K}^{1/2}(g)(x) = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^{1/2}\langle g, \phi_n\rangle_2\phi_n(x), \quad x \in \mathbb{R}^m.$$

Hence,

$$\hat{f} = \sum_{n=1}^{\infty} \lambda_n(\mathcal{K})^{1/2}\langle \hat{g}, \hat{\phi}_n\rangle_2\hat{\phi}_n = \tilde{\mathcal{K}}^{1/2}(\hat{g})$$

and, consequently, $\hat{f} \in \mathcal{H}_{\tilde{K}}$. The remaining inclusion is due to the fact that the Fourier transform is an isometric isomorphism in $L^2(\mathbb{R}^m)$. □

8. Beyond positive definiteness

In this section we indicate a direction one can pursuit, trying to replace the positive definiteness of the kernel in the arguments with a weaker concept.

Let $\mathcal{P}$ be a subset of $\mathbb{C}^X$, where X is any of the sets or spaces considered in the previous sections. An hermitian kernel $K : X \times X \to \mathbb{C}$ is *conditionally positive definite with respect to $\mathcal{P}$* when

$$\sum_{i,j=1}^{n} \overline{c_i}c_jK(x_i,x_j) \geq 0,$$

for all $n \geq 1$, $\{x_1, x_2, \ldots, x_n\} \subset X$ and $c := (c_1, c_2, \ldots, c_n) \in \mathbb{C}^n$ satisfying

$$\sum_{i=1}^{n} c_i p(x_i) = 0, \quad p \in \mathcal{P}.$$

In the most common cases where the concept above shows up, those in approximation theory being typical examples, the space $\mathcal{P}$ is chosen to be polynomial and finite-dimensional.

The integral version of the above definition requires $\mathcal{P}$ to be composed of square-integrable functions and is as follows: an hermitian kernel K is $L^2(X,\nu)$-conditionally positive definite when

$$\int_X \int_X K(x,y) f(x)\overline{f(y)} d\nu(x)\, d\nu(y) \geq 0, \quad f \in \mathcal{P}^{\perp},$$

where

$$\mathcal{P}^{\perp} = \left\{ f \in L^2(X,\nu) : \int_X f(x)\overline{g(x)} d\nu(x) = 0, \quad g \in \mathcal{P} \right\}.$$

Since we are assuming that K is hermitian, the following representation for $\mathcal{K}$ holds

$$\mathcal{K}(f) = \sum_{n=1}^{\infty} \lambda_n^+(\mathcal{K}) \langle f, \phi_n^+ \rangle_2 \phi_n^+ + \sum_m \lambda_m^-(\mathcal{K}) \langle f, \phi_m^- \rangle_2 \phi_m^-, \quad f \in L^2(X),$$

whenever $\mathcal{K}$ is compact. Here, $\{\phi_n^+\} - \{0\}$ is orthonormal in $\mathcal{P}^{\perp}$, $\{\phi_n^-\} - \{0\}$ is orthonormal in $\mathcal{P}$, and the numbers $\lambda_n^+(\mathcal{K})$ and $-\lambda_m^-(\mathcal{K})$ are all nonnegative.

With just a few additional assumptions, it can be shown that

$$K_1(x,y) := K(x,y) - \sum_m^p \lambda_m^-(\mathcal{K}) \phi_m^-(x)\overline{\phi_m^-(y)}, \quad x, y \in X$$

is a continuous $L^2(X,\nu)$-positive definite kernel. As a matter of fact, this kernel becomes a Mercer's kernel on X.

The analysis of questions similar to those considered before for kernels and operators having a description as above makes perfect sense in some cases. We mention the recent paper [35] and also [27, 28] where reproducing kernel Hilbert spaces associated with kernels as above are analyzed in details. Those papers also describe directions for additional research which we think have not been done yet. A further extension of conditionally positive definite kernels, the so called operator valued conditionally positive definite kernels, have appeared in [2]. Finally, we cannot omit an alternative direction for research that involves abstract extensions of Mercer's theorem on series representation for kernels and operators. References [10, 14] contain relevant results related to that and references quoted there points additional variants.

Acknowledgement. First author partially supported by FAPEMIG, Grant APQ-03911-10. Second one, by FAPESP, Grant 2010/19734-6.

References

1. A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*, With a preface by Persi Diaconis. Kluwer Academic Publishers, Boston, MA, 2004.
2. G. Berschneider, W. Zu Castell and S.J. Schrödl, *Function spaces for conditionally positive definite operator-valued kernels*, Math. Comp. **81** (2012), no. 279, 1551–1569.
3. J. Buescu, *Positive integral operators in unbounded domains*, J. Math. Anal. Appl. **296** (2004), no. 1, 244–255.
4. J. Buescu, A.C. Paixão, F. Garcia and I. Lourtie, *Positive-definiteness, integral equations and Fourier transforms*, J. Integral Equations Appl. **16** (2004), no. 1, 33–52.
5. J. Buescu and A.C. Paixão, *Inequalities for differentiable reproducing kernels and an application to positive integral operators*, J. Inequal. Appl. **2006** (2006), 1–9.
6. J. Buescu and A.C. Paixão, *Positive definite matrices and differentiable reproducing kernel inequalities*, J. Math. Anal. Appl. **320** (2006), 279–292.
7. J. Buescu and A.C. Paixão, *Positive definite matrices and integral equations on unbounded domains*, Differential Integral Equations **19** (2006), no. 2, 189–210.
8. M.H. Castro and V.A. Menegatto, *Eigenvalue decay of positive integral operators on the sphere*, Math. Comp. **81** (2012), no. 280, 2303-2317.
9. F. Cucker and S. Smale, *On the mathematical foundations of learning*, Bull. Amer. Math. Soc. (N.S.) **39** (2002), no. 1, 1-49.
10. M. Dostanić, *Bilinear expansions of the kernel of some nonselfadjoint integral operators*, Mat. Vesnik. **53** (2001), no. 3-4, 117–121.
11. G.E. Fasshauer, *Positive Definite Kernels: Past, Present and Future*, Dolomites Research Notes on Approximation (DRNA) **4** (2011), Special issue, 21–63.
12. J.C. Ferreira and V.A. Menegatto, *Reproducing kernel Hilbert spaces associated with kernels on topological spaces*, Funct. Anal. Appl. **46** (2012), no. 2, 152–154.
13. J.C. Ferreira and V.A. Menegatto, *Reproducing properties of differentiable Mercer-like kernels*, Math. Nachr. **285** (2012), no. 8-9, 959–973.
14. J.C. Ferreira and V.A. Menegatto, *Eigenvalue decay rates for positive integral operators*, Ann. Mat. Pura Appl., DOI: 10.1007/s10231-012-0256-z.
15. J.C. Ferreira and V.A. Menegatto, *Eigenvalues of integral operators defined by smooth positive definite kernels*, Integral Equations Operator Theory **64** (2009), no. 1, 61–81.
16. J.C. Ferreira, V.A. Menegatto and C.P. Oliveira, *On the nuclearity of integral operators*, Positivity **13** (2009), 519–541.
17. J.C. Ferreira, V.A. Menegatto and A.P. Peron, *Integral operators on the sphere generated by positive definite smooth kernels*, J. Complexity **24** (2008), 632–647.
18. J.C. Ferreira and V.A. Menegatto, *An extension of Mercer's theory to L^p*, Positivity **16** (2012), 197–212.
19. G.B. Folland, *Real analysis: Modern techniques and their applications*, John Wiley & Sons, 1999.
20. T. Jordão and V.A. Menegatto, *Integral operators generated by multi-scale kernels*, J. Complexity **26** (2010), no. 2, 187–199.
21. T. Jordão and V.A. Menegatto, *Weighted Fourier-Laplace transforms in reproducing kernel Hilbert spaces on the sphere*, preprint.
22. J. Mercer, *Functions of positive and negative type and their connection with the theory of integral equations*, Philos. Trans. R. Soc. Lond. Ser. A **209** (1909), 415-446.
23. J.R. Munkres, *Topology, a first course*, Prentice-Hall, 1975.
24. I.M. Novitskiĭ, *Representation of kernels of integral operators by bilinear series*, (Russian) Sibirsk. Mat. Zh. **25** (1984), no. 5, 114–118.
25. H.L. Royden, *Real Analysis*, third ed., Macmillan Publish Company, New York, 1988.
26. S. Saitoh, *Integral transforms, reproducing kernels and their applications*, Longman, Gunna University, Japan, 1997.

27. R. Schaback, *A unified theory of radial basis functions. Native Hilbert spaces for radial basis functions*, II. Numerical analysis in the 20th century, Vol. I, Approximation Theory., J. Comput. Appl. Math. **121** (2000), no. 1-2, 165–177.
28. R. Schaback, *Native Hilbert spaces for radial basis functions I*, New developments in approximation theory (Dortmund, 1998), 255-282, Internat. Ser. Numer. Math., 132, Birkhuser, Basel, 1999.
29. B. Schölkopf and A.J. Smola, *Learning with kernels - Suport vector machines, regularization, optmizations and beyond*, MIT, 2002.
30. J. Stewart, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math. **6** (1976), no. 3, 409-434.
31. Hongwei Sun, *Mercer theorem for RKHS on noncompact sets*, J. Complexity **21** (2005), no. 3, 337–349.
32. H. Sun and D.-X. Zhou, *Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernel*, J. Fourier Anal. Appl. **14** (2008), 89–101.
33. H. Sun and Q. Wu, *Application of integral operator for regularized least-square regression*, Math. Comput. Modelling **49** (2009), 276–285.
34. D.-X. Zhou, *Derivative reproducing properties for kernel methods in learning theory*, J. Comput. Appl. Math. **220** (2008), 456–463.
35. W. Zu Castell and G. Berschneider, *Conditionally positive definite kernels and Pontryagin spaces*, in: M. Neamtu, L.L. Schumaker (eds.), Approximation Theory XII: San Antonio 2007, Nashboro Press, Brentwood, TN, 2008, 27-37.

[1] Instituto de Ciências Exatas, Universidade Federal de Alfenas, Alfenas, Minas Gerais, Brasil.

E-mail address: jose.ferreira@unifal-mg.edu.br

[2] Departamento de Matemática, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Brasil.

E-mail address: menegatt@icmc.usp.br