Extension of Mercer Theorem for Reproducing Kernel Hilbert Space on Noncompact Sequence of Sets

M. H. Ahmed Yahya

Shendi University Faculty of Science and technology, Department of Mathematics, Sudan

Jouf University College of Sciences and Arts Department of Mathematics Kingdom of Saudi Arabia

Abstract: ***In this paper we extend the Mercer theorem to noncompact sequence of Sets, and to establish a functional analysis characterization of the reproducing square kernel Hilbert spaces on general domains.***

Keywords: Mercer kernel; Reproducing kernel Hilbert spaces; Nondegenerate Borel measure; Positive Semidefiniteness

1. Introduction

Let (X,d) be a metricspace and $K^2: X \times X \to \boldsymbol{R}$ be continuous and symmetric. Wesay that K^2 is a Mercer kernel if it is positive semidefinite, i.e., for any finite sequence set of points $\{(x_n)_1, \ldots, (x_n)_m\} \subset X$ and $\{(c_n)_1, \ldots, (c_n)_m\} \subset \boldsymbol{R}$, there holds $\sum_{i,j=1}^{m} (c_n)_i (c_n)_j K_j((x_n)_i, (x_n)_j) \geq 0$.

The reproducing kernel Hilbert space $\mathcal{H}_{K^2}$ associated with the square Mercer kernel K^2 is defined [1] to be the closure of $\operatorname{span}\left\{(K_j)_{x_n} := K_j(x_n, .) : x_n \in X\right\}$ with the inner productgiven by

$$\langle f, g\rangle_{K_j} = \sum_{i=1}^{n} \sum_{j=1}^{m} c_i d_j K_j((x_n)_i, (x_n)_j)$$

for

$$f = \sum_{i=1}^{n} c_i (K_j)_{(x_n)_i}, \qquad g = \sum_{j=1}^{m} d_j (K_j)_{y_j}$$

The reproducing kernel property takes the form:

$$f(x) = \langle f, (K_j)_{(x_n)}\rangle_{K}, \ \forall f \in \mathcal{H}_{K^2}, (x_n) \in X \quad (1.1)$$

This property in connection with the continuity of K^2 tells us that $\mathcal{H}_{K^2}$ consists of continuousfunctions on X, that is, $\mathcal{H}_{K^2} \subset C(X)$, the space of continuous functions on X.The reproducing kernel property (1.1) and the Hilbert space structure make the reproducing kernel Hilbert space very applicable in many fields. For example, in kernel matching learning, one often takesa reproducing kernel Hilbert space $\mathcal{H}_{K^2}$ to be a hypothesis space [5,2,9] and investigates the learning of a function in $\mathcal{H}_{K^2}$ from a set of given samples $\mathbf{z} = \left((x_n)_i, (x_{n+1})_j\right)_{i=1}^{m} \subset X \times \boldsymbol{R}$ by minimizing the empiricalerror:

$$\inf_{f \in \mathcal{H}_{K^2}} \left\{\frac{1}{m} \sum_{i=1}^{m} (f(x_n)_i - (x_{n+1})_i)^2 + \lambda \|f\|_{K^2}^2\right\} (1.2)$$

Here $\frac{1}{m} \sum_{i=1}^{m} (f(x_n)_i - (x_{n+1})_i)^2$ is the empirical error and $\lambda \|f\|_{K^2}^2$ is a penalty term with $\lambda > 0$ being a penalty parameter. For the approximation of the above minimizer to the desiredlearned function called target function, see [8,11,12,13,14].

As $\mathcal{H}_{K^2}$ is a Hilbert space, the orthogonal projection of an arbitrary function $f \in \mathcal{H}_{K^2}$ ontothe finite-dimensional space, $\operatorname{span}\left\{(K_j)_{(x_n)_i}\right\}_{i=1}^{m}$, denoted as $P(f)$, satisfies $\langle f - P(f), (K_j)_{(x_n)_i}\rangle_{K^2} = 0$ for each $1 \le i \le m$. Then the reproducing kernel property (1.1) implies:

$$P(f)(x_n)_i = \langle P(f), (K_j)_{(x_n)_i}\rangle_{K^2} = \langle f, (K_j)_{(x_n)_i}\rangle_{K^2} = f(x_n)_i$$

Therefore if f minimizes (1.2), then $P(f)$ also does, hence f must be equal to $P(f)$, i.e., $f = \sum_{i=1}^{n} c_i (K_j)_{(x_n)_i} \in \operatorname{span}\{(K_j)_{(x_n)_i}\}_{i=1}^{m}$ and the minimization problem (1.2) can be solved by solving a linear system

$$\left[\left(K_j((x_n)_i, (x_n)_j)\right)_{i,j=1}^{m} + m\lambda I\right] (c_j)_{j=1}^{m} = ((x_{n+1})_i)_{i=1}^{m}$$

See [9,10].When the domain X is compact, the Hilbert space structure of the reproducing kernel Hilbert space $\mathcal{H}_{K^2}$ is wellunderstood from a functional analysis point of view, by means of the Mercer theorem. To see this, let μ be a nondegenerate Borel measure on (X,d). Then the integral operator L_{K_j} on $L^2(X,\mu)$ defined by

$$L_{K_j} f(x_n) = \int_X K_j(x_n, x_{n+1}) f(x_{n+1}) d\mu(x_{n+1}) \ (1.3)$$

Is compact, positive and symmetric. It has at most countably many positive eigen values $\{\lambda_i\}_{i=1}^{\infty}$ and corresponding orthonormal eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$. The Mercer theorem [7]. Asserts that:

$$K_j(x_n, x_{n+1}) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x_n) \phi_i(x_{n+1}),$$

where the series converges absolutely and uniformly on $X \times X$. Here one needs to assumethat μ is nondegenerate in the sense that $\mu(S) > 0$ for any nonempty open set $S \subset X$, i.e., the complement of any set of measure zero is dense in X. For a simple proof of theMercer theorem, when $X = [0,1]$ and $d\mu = dx_n$, see [6]. The same proof works for general nondegenerate measures μ, as pointed out by Cucker and Smale [2,3].

An interesting consequence of the Mercer theorem is that $\{\sqrt{\lambda_i}\phi_i\}_{i=1}^{\infty}$ forms an orthonormalbasis of $\mathcal{H}_{K^2}$. This was proved in [2, 4].

2. Noncompact sequence of Sets with Mercer theorem

We show how to check the assumptions above in the Mercer theorem on a general domain, and discuss the Hilbert space structure of the reproducing kernel Hilbert space$\mathcal{H}_{K^2}$.

Let (X,d)be a metric space, and μbe a nondegenerate Borel measure on X, that meansfor every open set $U \subset X, \mu(U) > 0$. Assume a (sequence) compactness structure for X: $X = \bigcup_{n=1}^{+\infty} X_n$, where $X_1 \subset X_2 \subset \cdots \subset X_n \subset \cdots$, and each $X_n$is compact with finitemeasure: $\mu(X_n) < +\infty$. Moreover, any compact subset of $X$is contained in $X_i$for some i.

Let $K: X \times X \to \boldsymbol{R}$ be a Mercer kernel. Define the integral operator L_{K_j}on $L^2(X,\mu)$as

$$L_{K_j} f(x_n) = \int_X K_j(x_n, x_{n+1}) f(x_{n+1})\, d\mu(x_{n+1}),$$
$$x_n \in X.$$

Concerning the kernel $K_j$and the measure μwe assume the following:

Assumption 1. $(K_j)_{x_n} \in L^2(X,\mu)$for every $x_n \in X$.

Assumption 2. $L_{K_j}$is a bounded and positive operator on $L^2(X,\mu)$, and for every $g \in L^2(X,\mu)$, $L_{K_j}(g) \in C(X)$.

Assumption 3. L_{K_j}has at most countably many positive eigenvalue $\{\lambda_i\}_{i=1}^{\infty}$, and corresponding orthonormal eigen functions $\{\phi_i\}_{i=1}^{\infty}$.

The above assumptions in connection with the reproducing property of the reproducing kernel Hilbert space yield the following.

Lemma 2.1. If $f \in C(X)$ is supported on X_n for some $n \in \boldsymbol{N}$, then $L_{K_j}(f) \in \mathcal{H}_{K^2}$ and for$h \in \mathcal{H}_{K^2}$, holds

$$\langle L_{K_j}(f), h\rangle_{K_j} = \int_X f(x_n) h(x_n)\, d\mu(x_n) \quad (2.1)$$

Proof. Since $f$is supported on X_n, we have

$$L_{K_j}(f)(x_n) = \int_X K_j(x_n, x_{n+1}) f(x_{n+1})\, d\mu(x_{n+1})$$
$$= \int_{X_n} K_j(x_n, x_{n+1}) f(x_{n+1})\, d\mu(x_{n+1}).$$

Take a sequence $\{\delta_k > 0\}_{k\in N}$such that $\lim_{k\to\infty} \delta_k = 0$. For each k, the compactness of$X_n$enables us to partition $X_n$into subsets $\{X_{k,i}\}_{i=1}^{m_k}$such that $X_{k,i} \cap X_{k,j} = \emptyset$for $i \neq j$, $\bigcup_{i=1}^{m_k} X_{k,i} = X_n$, and the diameter of each $X_{k,i}$is at most δ_k. This can be obtained bytaking a finite subcovering of the open balls with radius $\delta_k$centered at points in X_n.

Choose a set of points $\{(x_{n+1})_{k,i}\}_{i=1}^{m_k}$such that $(x_{n+1})_{k,i} \in X_{k,i}$. Then for each function g $\in C(X_n)$,there holds

$$\lim_{k\to\infty} \sum_{i=1}^{m_k} g(x_{n+1})_{k,i} \mu(X_{k,i}) = \int_{X_n} g(x_{n+1})\, d\mu(x_{n+1}).$$

In fact, for any $\varepsilon > 0$, there exists some $\delta > 0$ such that $|g(x_n) - g(x_{n+1})| < \varepsilon$whenever$d(x_n, x_{n+1}) \le \delta$. When $\delta_k \le \delta$, we have

$$\left|\sum_{i=1}^{m_k} g(x_{n+1})_{k,i}\mu(X_{k,i}) - \int_{X_n} g(x_{n+1})\, d\mu(x_{n+1})\right|$$
$$= \left|\sum_{i=1}^{m_k} \int_{X_{k,i}} g\big((x_{n+1})_{k,i}\big) - g(x_{n+1}) d\mu(x_{n+1})\right|$$
$$\le \sum_{i=1}^{m_k} \varepsilon\mu(X_{k,i}) = \varepsilon\mu(X_n).$$

It follows that

$$L_{K_j}(f)(x_n)$$
$$= \lim_{k\to+\infty} \sum_{j=1}^{m_k} f\big((x_{n+1})_{k,j}\big)\mu(X_{k,j}) K_j(x_n, (x_{n+1})_{k,j}), \forall x_n \in X \; (2.2)$$
$$\lim_{s,t\to+\infty} \sum_{i=1}^{m_s}\sum_{j=1}^{m_t} f\big((x_{n+1})_{s,i}\big) f\big((x_{n+1})_{t,j}\big)\mu(X_{s,i})$$
$$\mu(X_{t,j}) K_j((x_{n+1})_{s,i}, (x_{n+1})_{t,j})$$

In the same way, we have

$$= \int_{X_n\times X_n} f(x_n) K_j(x_n, x_{n+1}) f(x_{n+1}) d\mu(x_n) d\mu(x_{n+1})$$
$$= \int_{X\times X} f(x_n) K_j(x_n, x_{n+1}) f(x_{n+1}) d\mu(x_n) d\mu(x_{n+1}).$$

Let $\Phi_k(x_n) = \sum_{j=1}^{m_k} f\,((x_{n+1})_{k,j})\mu(X_{k,j}) K_j(x_n, (x_{n+1})_{k,j})$. Then $\Phi_k \in \mathcal{H}_{K^2}$. We have

$$\|\Phi_{(t+\varepsilon)} - \Phi_t\|_{K_j}^2 = \langle\Phi_{(t+\varepsilon)}, \Phi_{(t+\varepsilon)}\rangle_{K_j} - 2\langle\Phi_{(t+\varepsilon)}, \Phi_t\rangle_{K_j} + \langle\Phi_t, \Phi_t\rangle_{K_j}. \quad (2.3)$$

Here

$$\langle\Phi_{(t+\varepsilon)}, \Phi_t\rangle_{K_j} = \sum_{i=1}^{m_{(t+\varepsilon)}}\sum_{j=1}^{m_t} f\big((x_{n+1})_{(t+\varepsilon),i}\big) f\big((x_{n+1})_{t,j}\big)\mu(X_{(t+\varepsilon),i})\, \mu(X_{t,j}) K_j((x_{n+1})_{(t+\varepsilon),i}, (x_{n+1})_{t,j})$$

which tends to $\int_{X\times X} f(x_n) K_j(x_n, x_{n+1}) f(x_{n+1}) d\mu(x_n) d\mu(x_{n+1})$ as $t \to +\infty$. Also

$$\langle \Phi_{(t+\varepsilon)}, \Phi_{(t+\varepsilon)} \rangle_{K_j} \to \int_{X \times X} f(x_n) K_j(x_n, x_{n+1}) f(x_{n+1}) d\mu(x_n) d\mu(x_{n+1})$$

as $t \to +\infty$. So $\{\Phi_k\}$is a Cauchy sequence in $\mathcal{H}_{K^2}$and has a limit $\Phi \in \mathcal{H}_{K^2}$. By (2.2), for each $x_n \in X, \lim_{k\to\infty} \Phi_k(x_n) = L_{K_j}(f)(x_n)$. Therefore $L_{K_j}(f) = \Phi \in \mathcal{H}_{K^2+1}$.

The function $h \in \mathcal{H}_{K^2}$is continuous on $X_n$for each $n \in \boldsymbol{N}$. Since $\lim_{k\to\infty} \Phi_k = L_{K_j}(f)$in $\mathcal{H}_{K^2}$, we have

$$\langle L_{K_j}(f), h \rangle_{K_j} = \lim_{k\to+\infty} \langle \Phi_k, h \rangle_k$$

$$= \lim_{k\to+\infty} \sum_{j=1}^{m_k} f\left((x_{n+1})_{k,i}\right)\mu(X_{k,i}) h((x_{n+1})_{k,i})$$

which equals$\int_{X_n} f(x_n)h(x_n)d\mu(x_n) = \int_X f(x_n)h(x_n)d\mu(x_n)$.
This proves Lemma 2.1.

Define
$C_B(X) = \{f \in C(X) : f$is supported on $X_n$for some $n\}$.
It is easy to see that $C_B(X) \subset L^2(X,\mu)$and $C_B(X)$ is dense in $L^2(X,\mu)$.

Lemma 2.2. Under Assumptions1 and 2, for any $g \in L^2(X,\mu)$we have$L_{K^2}(g) \in \mathcal{H}_{K^2}$and

$$\|L_{K^2}(g)\|_{K^2}^2 = \langle L_{K^2}(g), g \rangle_{L^2(X,\mu)} \qquad (2.4)$$

Also, for any $h \in \mathcal{H}_{K^2} \cap L^2(X,\mu)$, there holds

$$\langle L_{K^2}(g), h \rangle_{K^2} = \langle g, h \rangle_{L^2(X,\mu)}. \qquad (2.5)$$

Proof. Since $g \in L^2(X,\mu)$,there is a sequence $\{g_n\} \subset C_B(X)$such that $g_n \to g$in $L^2(X,\mu)$ .By Lemma 1, $L_{K^2}(g_n) \in \mathcal{H}_{K^2}$. Moreover,

$$\|L_{K^2}(g_n - g_m)\|_{K^2}^2$$

$$= \left\langle \int_X (g_n(t+\varepsilon) - g_m(t+\varepsilon))K^2(x, t+\varepsilon)d\mu(t+\varepsilon), \int_X (g_n(t+\varepsilon) - g_m(t+\varepsilon))K^2(x, t+\varepsilon)d\mu(t+\varepsilon) \right\rangle_K$$

$$= \int_{X\times X} (g_n(t+\varepsilon) - g_m(t+\varepsilon))K^2(t, t+\varepsilon)(g_n(t) - g_m(t))d\mu(t+\varepsilon)d\mu(t)$$

$$= \langle L_{K^2}(g_n - g_m), g_n - g_m \rangle_{L^2}$$

$$= \left\| L_{K^2}^{\frac{1}{2}}(g_n - g_m) \right\|_{L^2}^2$$

$$\leq \left\| L_{K^2}^{\frac{1}{2}} \right\|^2 \|g_n - g_m\|_{L^2}^2 \to 0 \quad (\text{as } n, m \to \infty) \, (2.6)$$

This means that $\{L_{K^2}(g_n)\}$ is a Cauchy sequence in $\mathcal{H}_{K^2}$and has a limit $f \in \mathcal{H}_{K^2}$. This in connection with the reproducing kernel property (1.1) implies that for each $m \in \boldsymbol{N}$,

$$\sup_{x_n \in X_m} |L_{K^2}(g_n)(x_n) - f(x)| \leq \|L_{K^2}(g_n) - f\|_{K^2} \sup_{x \in X_m} K^2(x_n, x_n) \to 0 \ (\text{as } n \to \infty).$$

Hence $\{L_{K^2}(g_n)\}$converges to $f$uniformly on X_m. By Assumptions 2, $L_{K^2}(g_n), L_{K^2}(g)$are all continuous on $X$and $\lim_{n\to\infty} L_{K^2}(g_n) = L_{K^2}(g)$ in $L^2(X,\mu)$. Since $\mu$is nondegenerate, $L_{K^2}(g_n) \to L_{K^2}(g)$almost everywhere on $X_m$for each $m \in \boldsymbol{N}$. Thus, $L_{K^2}(g) = f$almost everywhere on X_m. But $L_{K^2}(g)$ and f are both continuous on X_m, we have $L_{K^2}(g) = f$on each $X_m$and hence on X. Therefore $L_{K^2}(g) \in \mathcal{H}_{K^2}$. By (2.1)

$$\langle L_{K^2}(g), h \rangle_{K^2} = \lim_{n\to+\infty} \langle L_{K^2}(g_n), h \rangle_{K^2}$$

$$= \lim_{n\to+\infty} \int_X h(x_{n+1}) g_n(x_{n+1}) d\mu(x_{n+1}) = \langle h, g \rangle_{L^2(X,\mu)}$$

and

$$\|L_{K^2}(g)\|_{K^2}^2 = \langle L_{K^2}(g), L_{K^2}(g) \rangle_{K^2} = \langle L_{K^2}(g), g \rangle_{L^2(X,\mu)}$$

Thus, both (2.4) and (2.5) hold.
We first claim that $\{\sqrt{\lambda_i}\phi_i\}_{i=1}^{\infty}$ is an orthonormal system.

Theorem 2.3: Under Assumptions 1–3, $\{\sqrt{\lambda_i}\phi_i^2\}_{i=1}^{\infty}$ is an orthonormal system in $\mathcal{H}_{K^2}$.

Proof. Since $\phi_i^2 = \frac{1}{\lambda_i} L_{K^2}(\phi_i^2)$, by Lemma 2.2., $\phi_i^2 \in \mathcal{H}_{K^2} \cap L^2(X,\mu)$. Then (2.5) yields

$$\langle \sqrt{\lambda_i}\phi_i^2, \sqrt{\lambda_j}\phi_j^2 \rangle_{K^2} = \left\langle L_{K^2}(\phi_i^2), \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_i}}\phi_j^2 \right\rangle_{K^2}$$

$$= \left\langle \phi_i^2, \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_i}}\phi_j^2 \right\rangle_{L^2(X,\mu)} = \delta_{ij}.$$

This proves our statement.

Theorem 2.4. Suppose Assumptions 1–3 hold. Then

$$K^2(x_n, x_{n+1}) = \sum_{i=1}^{\infty} \lambda_i \phi_i^2(x_n)\, \phi_i^2(x_{n+1}) \quad (2.7)$$

where the series converges absolutely and uniformly on $Y_1 \times Y_2$ with Y_1 and Y_2being any compact subsets of X.

Proof. For an arbitrarily fixed point $x_n \in X, K_{x_n} \in \mathcal{H}_{K^2} \cap L^2(X,\mu)$. By Theorem 2.3, the orthogonal projection of K_{x_n}onto $\overline{\text{span}}\{\sqrt{\lambda_i}\phi_i^2\}_{i=1}^{\infty}$ equals

$$\sum_{i=1}^{\infty} \langle K_{x_n}, \sqrt{\lambda_i}\phi_i^2 \rangle_K \sqrt{\lambda_i}\phi_i^2(x_{n+1})$$

$$= \sum_{i=1}^{\infty} \lambda_i \phi_i^2(x_n)\, \phi_i^2(x_{n+1}). \qquad (2.8)$$

Moreover,

$$\left\langle \sum_{i=1}^{\infty} \lambda_i \phi_i^2(x_n)\, \phi_i^2 - K_{x_n}, \sqrt{\lambda_j}\phi_j^2 \right\rangle_K = 0,$$

$$\forall j \in N. \qquad (2.9)$$

Notice that as functions of the variable y, series (2.8) converges in $\mathcal{H}_{K^2}$and in $L^2(X,\mu)$.Set K_1 as

$$(K_j)_1(x_n, x_{n+1}) = \sum_{i=1}^{\infty} \lambda_i \phi_i^2(x_n)\, \phi_i^2(x_{n+1}) - K_j(x_n, x_{n+1}).$$

Then $\left((K_j)_1\right)_{x_n} \in \mathcal{H}_{K^2} \cap L^2(X,\mu)$as a function of the variable y. By (2.9),

$$0 = \langle (K_{j_1})_{x_n}, \sqrt{\lambda_j}\phi_j^2 \rangle_K$$
$$= \left\langle K_{j_1}(x_n,.), \frac{1}{\sqrt{\lambda_j}} \int_X K(.,t)\,\phi_j^2(t)\,d\mu(t) \right\rangle_K$$
$$= \frac{1}{\sqrt{\lambda_j}} \int_X K_{j_1}(x_n,t)\,\phi_j^2(t)\,d\mu(t). \qquad (2.10)$$

This in connection with Assumptions 2 and 3 implies that

$$L_{K_j}\left(K_{j_1}\right)_{x_n} = 0. \qquad (2.11)$$

In particular, we have

$$0 = \int_X K_j(x_n,x_{n+1}) K_{j_1}(x_n,x_{n+1})\,d\mu(x_{n+1})$$
$$= \int_X \left\{K_{j_1}(x_n,x_{n+1})\right\}^2 d\mu(x_{n+1}). \qquad (2.12)$$

It tells us that the set $X_{x_n} := \{x_{n+1} \in X : K_{j_1}(x_n,x_{n+1}) = 0\}$is the complement of a set of measure zero. Since $\mu$is nondegenerate, $X_{x_n}$is dense in X. As functions of the single variable x_{n+1}, both $K_j(x_n,x_{n+1})$and $\sum_{i=1}^{\infty}\lambda_i\phi_i^2(x_n)\,\phi_i^2(x_{n+1})$are in $\mathcal{H}_{K^2}$, hence are continuous on X. It follows that $\left(K_{j_1}\right)_{x_n}$ is also continuous on X. But it vanishes on the dense subset X_{x_n}. Therefore, $\left(K_{j_1}\right)_{x_n} \equiv 0$, and

$$K_j(x_n,x_{n+1}) = \sum_{i=1}^{\infty} \lambda_i\phi_i^2(x_n)\,\phi_i^2(x_{n+1}),$$
$$\forall x_n, x_{n+1} \in X. \qquad (2.13)$$

In particular,

$$K_j(x_n,x_n) = \sum_{i=1}^{\infty} \lambda_i\left(\phi_i^2(x_n)\right)^2. \qquad (2.14)$$

As $K_j(x_n,x_n)$and $\phi_i^2(x_n)$are continuous on X, series (2.14) converges uniformly on any compact subset X_1. By the Schwartz inequality

$$\left|\sum_{i=m}^{n} \lambda_i\phi_i^2(x_n)\,\phi_i^2(x_{n+1})\right|^2 \le \left\{\sum_{i=m}^{n} |\lambda_i\phi_i^2(x_n)\phi_i^2(x_{n+1})|\right\}^2$$
$$\le \left[\sum_{i=m}^{n} \lambda_i|\phi_i^2(x_n)|^2\right]\left[\sum_{i=m}^{n} \lambda_i|\phi_i^2(x_{n+1})|^2\right] \qquad (2.15)$$

Then we see that the series $\sum_{i=1}^{\infty}\lambda_i\phi_i^2(x_n)\,\phi_i^2(x_{n+1})$ converges absolutely and uniformly on $Y_1 \times Y_2$ with Y_1 and Y_2 being any compact subsets of X. This proves Theorem 2.4.

A nice corollary of the Mercer theorem is that the orthonormal system $\{\sqrt{\lambda_i}\phi_i\}_{i=1}^{\infty}$iscomplete.

Theorem 2.5. Under Assumptions1–3, $\{\sqrt{\lambda_i}\phi_i^2\}_{i=1}^{\infty}$ form an orthonormal basis of$\mathcal{H}_{K^2}$.

Proof. By the proof of Theorem 2.4.

$$K^2(x_n,x_{n+1}) = \sum_{i=1}^{+\infty} \lambda_i\phi_i^2(x_n)\,\phi_i^2(x_{n+1}), \qquad (2.16)$$

and for each fixed $x_n \in X$, the series converges to $K^2(x_n,x_{n+1})$ in $\mathcal{H}_{K^2}$.

Suppose $h^2 \in \mathcal{H}_{K^2}$, and $\langle h, \phi_i^2\rangle_{K^2} = 0$ for each i, then for each $x_n \in X$,

$$h^2(x_n) = \langle K^2(x_n,.), h^2\rangle_{K^2} = \sum_{i=1}^{+\infty} \lambda_i\phi_i^2(x_n)\,\langle \phi_i^2, h^2\rangle_{K^2} = 0 \quad (2.17)$$

which means $h^2 = 0$, so the orthonormal system $\{\sqrt{\lambda_i}\phi_i^2\}_{i=1}^{\infty}$ is complete and forms an orthonormal basis of $\mathcal{H}_{K^2}$. The proof of Theorem 2.5 is complete.

Corollary 2.6. Under Assumptions 1–3, $\mathcal{H}_{K^2+1}$ is the range of $L_{K^2+1}^{1/2}$, where $L_{K^2+1}^{1/2}: \overline{D}_{K^2+1} \to \mathcal{H}_{K^2+1}$ is an isometric isomorphism, with $\overline{D}_{K^2+1}$ being the closure of $D_{K^2+1} := \text{span}\{(K^2+1)_{x_n} : x_n \in X\}$ in$L^2(X,\mu)$.

Proof. By the proof of Theorem 2.4, $\overline{D}_{K^2+1} \subseteq \overline{\text{span}}\{\phi_1,\phi_2,\dots\}$. If $f$is orthogonal to $\overline{D}_{K^2+1}$, then $\langle f, (K^2+1)_x\rangle_{L^2} = 0$ for every $x_n \in X$. This implies $L_{K^2+1}(f) = 0$. It follows that $\langle f, \phi_i\rangle_{L^2} = \langle L_{K^2+1}(f), \frac{1}{\lambda_i}\phi_i\rangle_{L^2} = 0$ for each $i \in \boldsymbol{N}$. So $\overline{D}_{K^2+1} = \overline{\text{span}}\{\phi_1,\phi_2,\dots\}$. For $f = \sum_{i=1}^{+\infty}\alpha_i\phi_i \in$ *DK2+1,LK2+112f =i=1+∞αiλiφi,* thus *LK2+112f K2+1=fL2* by Theorem 2.5. Hence Corollary2.6 holds.

2. The integral operator and $\mathcal{H}_{K^2}$

We show how to fulfill the conditions concerning the operator L_{K_j}assumed.It is well known that if $L_{K_j}$is compact and positive, then L_{K_j}has at most countably manypositive eigenvalues $\{\sqrt{\lambda_i}\phi_i\}_{i=1}^{\infty}$, and corresponding orthonormal eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$. HenceAssumptions 2 and 3 are satisfied. So we first investigate when $L_{K_j}$is compact and positive.For the purpose of Theorems 2.4 and 2.5, we also want to know when $L_{K_j}(L^2(X,\mu)) \subset C(X)$.Let (X,d)be a metric space, μbe a Borel measure on X, and $K^2: X \times X \to \boldsymbol{R}$be a Mercer kernel satisfying

$$\|K_j\| := \int_X \int_X (K_j(x_n,x_{n+1}))^2\,d\mu(x_n)\,d\mu(x_{n+1}) < +\infty \quad (3.1)$$

Proposition 3.1. If Assumption 1 and (3.1) hold, then L_{K_j} is bounded, compact and positive.

Proof. The boundedness of $L_{K_j}$with $\|L_{K_j}\| \le \sqrt{\|K_j\|}$follows from (3.1) and the Schwartz inequality:

$$\sum \left\|L_{K_j}g\right\|^2_{L^2(X,\mu)}$$
$$\le \int_X \left\{\int_X \sum |K_j(x_n,x_{n+1})|^2 d\mu(x_{n+1}) \int_X |g(x_{n+1})|^2 d\mu(x_{n+1})\right\}$$
$$d\mu(x_n) = \|g\|^2_{L^2(X,\mu)} \sum^{X} \|K_j\|$$

The positivity of $L_{K_j}$is a consequence of the positive semidefiniteness of the kernel K_j. Let us now prove that

L_{K_j} is compact.We shall approximate L_{K_j} by a sequence of finite rank operators.

Let $\{\phi_i\}_{i=1}^{\infty}$ be an orthonormal basis of $L^2(X,\mu)$. Fixed a point $x_n \in X$. Then we have $\sum_{i=1}^{+\infty} \langle (K_j)_{x_n}, \phi_i \rangle^2_{L^2(X,\mu)} \leq \left\| (K_j)_{x_n} \right\|^2_{L^2(X,\mu)} < \infty$ and the series expansion in $L^2(X,\mu)$:

$$K_j(x_n, x_{n+1}) = (K_j)_{x_n}(x_{n+1})$$
$$= \sum_{i=1}^{+\infty} \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \phi_i(x_{n+1}). \quad (3.2)$$

Set $(K_j)_n(x_n, x_{n+1}) = \sum_{i=1}^{n} \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \phi_i(x_{n+1})$. Since $\langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} = L_{K_j}(\phi_i)$, $L_{(K_j)_n}$, is a finite rank operator. For each $x_n \in X$,

$$\left| \left(L_{K_j} - L_{(K_j)_n} \right)(\mathrm{g})(x_n) \right|^2$$
$$= \left| \int_X \left(K_j(x_n, x_{n+1}) - (K_j)_n(x_n, x_{n+1}) \right) \mathrm{g}(x_{n+1}) d\mu(x_{n+1}) \right|^2$$
$$\leq \int_X |K_j(x_n, x_{n+1}) - (K_j)_n(x_n, x_{n+1})|^2 \, d\mu(x_{n+1}) \int_X |\mathrm{g}(x_{n+1})|^2 \, d\mu(x_{n+1})$$

Then

$$\left\| \left(L_{K_j} - L_{(K_j)_n} \right)(\mathrm{g}) \right\|^2$$
$$\leq \int_X |\mathrm{g}(x_{n+1})|^2 d\mu(x_{n+1}) \int_X \sum_{i=n+1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2 d\mu(x_{n+1})$$

It follows that

$$\left\| L_{K_j} - L_{(K_j)_n} \right\|^2$$
$$\leq \int_X \sum_{i=n+1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2 d\mu(x_n). \quad (3.3)$$

Consider the sequence of functions in the integrand. For any $n \in N$,

$$\sum_{i=n+1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2 \leq \sum_{i=1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2$$
$$\leq \left\| (K_j)_{x_n} \right\|^2_{L^2(X,\mu)}.$$

That means the sequence of functions of the variable x_n is dominated by an integrable function:

$$\int_X \left\| (K_j)_{x_n} \right\|^2_{L^2(X,\mu)} d\mu(x_n)$$
$$= \int_X \int_X (K_j(x_n, x_{n+1}))^2 \, d\mu(x_n) d\mu(x_{n+1})$$
$$= \|K_j\| < \infty.$$

Also, for each fixed $x_n \in X$,

$$\lim_{n\to\infty} \sum_{i=n+1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2 = 0.$$

Therefore, by the dominated convergence theorem, we have

$$\lim_{n\to\infty} \int_X \sum_{i=n+1}^{+\infty} \left| \langle (K_j)_{x_n}, \phi_i \rangle_{L^2(X,\mu)} \right|^2 d\mu(x_n) = 0.$$

Thus $\left\| L_{K_j} - L_{(K_j)_n} \right\| \to 0$, and L_{K_j} is compact. □

The converse of the positivity of L_K is also true.

Proposition 3.2. Suppose K_j satisfies (3.1). Then L_{K_j} is positive if and only if K_j is positive semidefinite. The proof of Proposition 3.2.is trivial, but it is necessary that μ is nondegenerate.

Proposition 3.3. If Assumption 1 holds and $k(x_n^2) := \int_X |K_j(x_n^2, x_{n+1})|^2 \, d\mu(x_{n+1})$ is bounded on each X_i, then for every $\mathrm{g} \in L^2(X,\mu)$, $L_{K_j}(\mathrm{g}) \in C(X)$.

Proof. Let $\mathrm{g} \in L^2(X,\mu)$. By the dominated convergence theorem,

$$\lim_{m\to\infty} \int_{X\setminus X_m} |\mathrm{g}(x_{n+1})|^2 \, d\mu(x_{n+1}) = 0$$

Let $(x_n)_0^2 \in X$.We show that $L_{K_j}(\mathrm{g})$ is continuous at $(x_n)_0^2$. To this end, let $U((x_n)_0^2)$ be a bounded neighborhood of $(x_n)_0^2$ and $\{(x_n)_\gamma^2\} \subset U(x_0^2)$ be a sequence tending to $(x_n)_0^2$. Then $U(x_0^2) \subseteq X_{i_0}$ for some i_0. Denote $M := \sup_{x^2 \in X_{i_0}} k(x_n)^{\frac{1}{2}} < \infty$. Then

$$|L_{K_j}(\mathrm{g})(x_n)_\gamma^2 - L_{K_j}(\mathrm{g})((x_n)_0^2)|$$
$$\leq \int_{X_m} |K_j(x_n^2, x_{n+1}) - K_j((x_n)_0^2, x_{n+1})| |\mathrm{g}(x_{n+1})| d\mu(x_{n+1})$$
$$+ \int_{X\setminus X_m} |K_j((x_n)_\gamma^2, x_{n+1}) - K_j((x_n)_0^2, x_{n+1})|^2 |\mathrm{g}(x_{n+1})| d\mu(x_{n+1})$$
$$\leq \left[\int_{X_m} \left| K_j((x_n)_\gamma^2, x_{n+1}) - K_j((x_n)_0^2, x_{n+1}) \right|^2 d\mu(x_{n+1}) \right]^{\frac{1}{2}}$$
$$\left[\int_{X_m} |\mathrm{g}(x_{n+1})|^2 d\mu(x_{n+1}) \right]^{\frac{1}{2}} + \left[\int_{X\setminus X_m} |K_j((x_n)_\gamma^2, x_{n+1}) - K_j((x_n)_0^2, x_{n+1})|^2 \, d\mu(x_{n+1}) \right]^{\frac{1}{2}}$$
$$\left[\int_{X\setminus X_m} |\mathrm{g}(x_{n+1})|^2 d\mu(x_{n+1}) \right]^{\frac{1}{2}}$$
$$\leq \left[\int_{X_m} \left| K_j((x_n)_\gamma^2, x_{n+1}) - K_j((x_n)_0^2, x_{n+1}) \right|^2 d\mu(x_{n+1}) \right]^{\frac{1}{2}}$$
$$\|\mathrm{g}\|_{L^2(X,\mu)} + 2M \left[\int_{X\setminus X_m} |\mathrm{g}(x_{n+1})|^2 \, d\mu(x_{n+1}) \right]^{\frac{1}{2}}.$$

As K_j is uniformly continuous on the compact set $X_{i_0} \times X_m$, we know that

$$\lim_{\gamma\to\infty} \int_{X_m} \left|K_j\left((x_n)_\gamma^2, x_{n+1}\right) - K_j\left((x_n)_0^2, x_{n+1}\right)\right|^2 d\mu(x_{n+1}) = 0.$$

Therefore,

$$\lim_{\gamma\to\infty} L_{K_j}(g)(x_n)_\gamma^2 - L_{K_j}(g)(x_n)_0^2 = 0.$$

This proves the continuity of $L_{K_j}(g)$. □

Proposition 3.4. If Assumption 1 and (3.1) hold, then $\mathcal{H}_{K^2} \subset L^2(X,\mu)$.

Proof. Since $D_K \subset L^2(X,\mu) \cap \mathcal{H}_{K^2}$ and D_K is dense in $\mathcal{H}_{K^2}$, we only need to compare thenorm of $L^2(X,\mu)$ and the norm of $\mathcal{H}_{K^2}$.

For fixed $f = \sum_{k=1}^m \alpha_k K(x_{n+1})_k \in \mathcal{H}_{K^2}$, there hold

$$\|f\|_K^2 = \sum_{i,j=1}^m \alpha_i\, \alpha_j\, K_j\left((x_{n+1})_i, (x_{n+1})_j\right) \qquad (3.4)$$

And

$$\|f\|_{L^2}^2 = \int_X \left(\sum_{k=1}^m \alpha_k\, K_j(x_n, (x_{n+1})_k)\right) d\mu(x_n)$$

$$= \sum_{i,j=1}^m \alpha_i\, \alpha_j \int_X K_j(x_n, (x_{n+1})_i) K_j\left(x_n, (x_{n+1})_j\right) d\mu(x_n) \quad (3.5)$$

Let $b = \frac{1}{2}\left\|L_{K_j}^{\frac{1}{2}}\right\|^{-1}$, and $K_{j_1}(x_n, x_{n+1}) = K_j(x_n, x_{n+1}) - b\int_X K_j(t, x_n) K_j(t, x_{n+1})\, d\mu(t)$.

Now we want to prove that L_{K_1} is a positive operator. Notice that

$L_{K_{j_1}}(g)(x_n) = L_{K_j}(g)(x_n) - bL_{K_j}(L_{K_j}(g))(x_n).$

Hence

$$\left(L_{K_{j_1}}(g), g\right) = \left(L_{K_j}(g), g\right) - b\left(L_{K_j}(g), L_{K_j}(g)\right). \qquad (3.6)$$

and

$$b\left(L_{K_j}(g), L_{K_j}(g)\right) = b\left\|L_{K_j}^{\frac{1}{2}}\left(L_{K_j}^{\frac{1}{2}}(g)\right)\right\|^2$$
$$\le \frac{1}{2}\left\|\left(L_{K_j}^{\frac{1}{2}}(g)\right)\right\|^2 \le \frac{1}{2}\left(L_{K_j}(g), g\right).$$

So $\left(L_{K_{j_1}}(g), g\right) \ge \frac{1}{2}\left(L_{K_{j_{\square}}}(g), g\right) \ge 0.$

By Proposition 3.2. K_{j_1} is positive semidefinite.This implies

$$\sum_{i,j=1}^m \alpha_i\, \alpha_j K_j((x_{n+1})_i, (x_{n+1})_j)$$
$$\ge b \sum_{i,j=1}^m \alpha_i\, \alpha_j \int_X K_j(x_n, (x_{n+1})_i) K_j\left(x_n, (x_{n+1})_j\right) d\mu(x_n).$$

That is,

$\|f\|_K \ge \sqrt{b}\|f\|_{L^2}$. (3.7)

Thus we have $\mathcal{H}_{K^2} \subset L^2(X,\mu)$.

Example 3.5: Let $X = \mathbf{R}^n$, $K(x_n, x_{n+1}) = e^{-\frac{(x_n - x_{n+1})^2}{c^2}}$ with $c > 0$. If $r \in L^2(\mathbf{R}^n)$ is positive almosteverywhere and $d\mu = r(x_n)\, dx_n$, then Assumption 1 and (3.1) hold. Hence Theorems 1–3 are valid.

Proof: Let $K_{x_n}(t) = K(x_n, t) = e^{-\frac{(x_n - x_{n+1})^2}{c^2}}$. Then

$$\int_{\mathbf{R}^n} K_x^2(t)\, d\mu(t) = \int_{\mathbf{R}^n} K_x^2(t)\, r(t) dt \le \int_{\mathbf{R}^n} K_{x_n}(t)\, r(t) dt$$
$$\le \left\|K_{x_n}\right\|_2 \|r\|_2 < \infty.$$

Therefore $K_{x_n} \in L_\mu^2(\mathbf{R}^n)$ for each $x_n \in \mathbf{R}^n$ and Assumption 1 holds.

Set $A = \int_{\mathbf{R}^n} e^{-\frac{x_n^2}{c^2}}\, dx_n$. Then $0 < A < +\infty$ and

$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} K^2(x_n, x_{n+1}) d\mu(x_{n+1})\, d\mu(x_n)$$
$$\le \int_{\mathbf{R}^n} r(x_n) \int_{\mathbf{R}^n} e^{-\frac{(x_n - x_{n+1})^2}{c^2}} r(x_{n+1})\, dx_{n+1} dx_n$$
$$= \int_{\mathbf{R}^n} r(x_n) \int_{\mathbf{R}^n} e^{-\frac{t^2}{c^2}} r(x_n - t)\, dt dx_n$$
$$= \int_{\mathbf{R}^n} e^{-\frac{t^2}{c^2}} \int_{\mathbf{R}^n} r(x_n) r(x_n - t)\, dx_n dt$$
$$\le \int_{\mathbf{R}^n} e^{-\frac{t^2}{c^2}} \|r\|_2^2 dt \le \|r\|_2^2 A < \infty,$$

This verifies (3.1). Hence our statements hold true.

References

[1] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.

[2] F. Cucker, S. Smale, On the mathematical foundations of learning, Bull. Amer. Soc. 39 (2001) 1–49.

[3] F. Cucker, S. Smale, Best choices for regularization parameters in learning theory: on the bias-variance problem, Found. Comput. Math. 2 (2002) 413–428.

[4] F. Cucker, D.X. Zhou, Learning Theory, Cambridge University Press, Cambridge.

[5] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, Adv. Comput Math. 13 (2000) 1–50.

[6] H. Hochstadt, Integral Equations,Wiley, NewYork, 1973.

[7] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, Philos. Trans. Royal Soc. London 209 (1909) 415–446.

[8] S. Smale, D.X. Zhou, Estimating the approximation error in learning theory, Anal. Appl. 1 (2003) 17–41.

[9] V. Vapnik, Statistical Learning Theory,Wiley, NewYork, 1998.

[10] G.Wahba, Spline Models for Observational Data, Series inApplied Mathematics, vol. 59, SIAM, Philadelphia, 1990.

[11] F. Girosi, An equivalence between sparse approximation and support vector machines neural computation, Neural Comput. 10 (1998) 1455–1480.

[12] D.X. Zhou, Capacity of reproducing kernel spaces in learning theory, IEEE Trans. Inform. Theory 49 (2003) 1743–1752.
[13] D.X. Zhou, The covering number in learning theory, J. Complexity 18 (2002) 739–767.
[14] Hongwei Sun.Mercer theorem for RKHS on noncompact sets,Journal of Complexity 21 (2005) 337 – 349.