On the speed of uniform convergence in Mercer’s theorem

RUSTEM TAKHANOV*, School of Sciences and Humanities, Republic of Kazakhstan

The classical Mercer’s theorem claims that a continuous positive definite kernel \( K(x, y) \) on a compact set can be represented as

\[
\sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(y)
\]

where \( (\lambda_i, \phi_i) \) are eigenvalue-eigenvector pairs of the corresponding integral operator. This infinite representation is known to converge uniformly to the kernel \( K \). We estimate the speed of this convergence in terms of the decay rate of eigenvalues and demonstrate that for 2m times differentiable kernels the first \( N \) terms approximate \( K \) as \( O((\sum_{i=N+1}^{\infty} \lambda_i^m)^{1/m}) \) or \( O((\sum_{i=N+1}^{\infty} \lambda_i^m)^{\infty/m}) \). Finally, we demonstrate some applications of our results to a spectral characterization of integral operators with continuous roots and other powers.

Additional Key Words and Phrases: Mercer’s theorem, Mercer kernel, uniform convergence, RKHS, Gagliardo-Nirenberg inequality.

ACM Reference Format:
Rustem Takhanov. 2022. On the speed of uniform convergence in Mercer’s theorem. 1, 1 (September 2022), 11 pages.

1 INTRODUCTION
Mecer kernels play an important role in machine learning and is a mathematical basis of such techniques as kernel density estimation and spline models [14], Support Vector Machines [11], kernel principal components analysis [10], regularization of neural networks [13] and many others. According to Aronszajn’s theorem, any Mercer kernel induces a reproducing kernel Hilbert space (RKHS) and vice versa, any RKHS corresponds to a kernel. A relationship between the latter two notions is described in the classical Mercer’s theorem. A goal of this note is to refine the theorem and give some estimates on the speed of uniform convergence stated in it.

Let \( \Omega \subseteq \mathbb{R}^n \) be a compact set, \( \Omega : \Omega \times \Omega \rightarrow \mathbb{R} \) be a continuous Mercer kernel [6] and \( L_p(\Omega) \), \( p \geq 1 \) be a space of real-valued functions \( f \) on \( \Omega \) with \( \|f\|_{L_p(\Omega)} = (\int_{\Omega} |f(x)|^p \, dx)^{1/p} \). Let \( O_K : L_2(\Omega) \rightarrow L_2(\Omega) \) be defined by \( O_K(\phi)(x) = \int_{\Omega} K(x, y) \phi(y) \, dy \). By \( C(\Omega) \) we denote a space of continuous functions. From Mercer’s theorem we have that there is an orthonormal basis \( \{\psi_i(x)\}_{i=0}^{\infty} \) in \( L_2(\Omega) \) such that \( O_K[\psi_i] = \lambda_i \psi_i \). Some of eigenvalues of \( O_K \) can be equal to zero, therefore, let us assume that natural numbers \( i_1 < i_2 < \cdots \) are such that \( \{\lambda_{i_j}\}_{j=1}^{\infty} \) is a set of positive eigenvalues, and we denote \( \lambda_i = \lambda'_{i_j} \) and \( \phi_j = \psi_{i_j} \), \( j \in \mathbb{N} \). It is well-known that \( \{\phi_i(x)\}_{i=0}^{\infty} \subseteq C(\Omega) \) and \( L_2^{\infty} = \|K(x, y) - \sum_{i=1}^{N} \lambda_i \phi_i(x) \phi_i(y)\|_{L_2(\Omega \times \Omega)}^2 = \sum_{i=N+1}^{\infty} \lambda_i^2 \). Analogously, for diagonal elements we have

\[
S_N = \|K(x, x) - \sum_{i=1}^{N} \lambda_i \phi_i(x)^2\|_{L_2(\Omega)} = \sum_{i=N+1}^{\infty} \lambda_i.
\]

Thus, the behaviour of eigenvalues completely characterizes the speed of convergence of \( \sum_{i=1}^{N} \lambda_i \phi_i(x) \phi_i(y) \) to \( K \) in \( L_2(\Omega \times \Omega) \) and of \( \sum_{i=1}^{N} \lambda_i \phi_i(x)^2 \) to \( K(x, x) \) in \( L_2(\Omega) \). For the supremum norm, Mercer’s theorem implies only the uniform convergence, i.e.

\[
C_N = \sup_{x, y \in \Omega} |K(x, y) - \sum_{i=1}^{N} \lambda_i \phi_i(x) \phi_i(y)| \rightarrow 0
\]
as \( N \rightarrow \infty \). We are interested in upper bounds on \( C_N \).

Author’s address: Rustem Takhanov, rustem.takhanov@nu.edu.kz, School of Sciences and Humanities, 53 Kabanbay Batyr Ave, Nur-Sultan city, Republic of Kazakhstan, 010000.

© 2022
For \( \alpha = (\alpha_1, \cdots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n \), \( |\alpha| \) denotes \( \sum_{i=1}^n \alpha_i \). \( \partial^\alpha f(x) \) denotes \( \frac{\partial^{\alpha_i} f(x)}{\partial x_i^{\alpha_i}} \). The symbol \( \mathcal{C}^m(\Omega) \) denotes a set of functions \( f : \Omega \to \mathbb{R} \) such that \( \partial^\alpha f \in \mathcal{C}(\Omega) \) for \( |\alpha| \leq m \). We prove the following theorems.

**Theorem 1.1.** Let \( \Omega \) have a Lipschitz boundary, \( K \in C^{2m}(\Omega \times \Omega) \) and \( p > \frac{n}{m} \). \( \rho \geq 1 \). Then,

\[
\mathcal{C}^2_K \leq C_{\Omega, p}^m(\max_{|\alpha|=m} \left( \sum_{\beta \leq \alpha} \theta \right) \|D^\beta \partial_{\alpha-\beta}f\|_{L^p(\Omega)}^\theta \left( \sum_{i=N+1}^\infty \lambda_i \right)^{-1}) + C_{\Omega, p}^m \sum_{i=N+1}^\infty \lambda_i
\]

where \( D_\alpha(x) = \partial_\alpha^\alpha \partial_\beta^\beta K(x, y) \). \( \theta = (1 + \frac{m}{n} - \frac{1}{p})^{-1} \) and

\[
C_{\Omega, p}^m = \sup_{u \in L^1(\Omega)} \frac{\|u\|_{L^1(\Omega)}}{\|u\|_{L^p(\Omega)}}
\]

is an optimal constant in the Gagliardo-Nirenberg inequality for the domain \( \Omega \).

Note that in the latter theorem one can set \( p = +\infty \) and obtain that \( \mathcal{C}^2_K = O((\sum_{i=N+1}^\infty \lambda_i)^{\frac{m}{n}}) \). Thus, infinitely differentiable kernels satisfy \( \mathcal{C}^2_K = O((\sum_{i=N+1}^\infty \lambda_i)^{1-\epsilon}) \) for any \( \epsilon > 0 \).

**Theorem 1.2.** Let \( \Omega \) have a Lipschitz boundary, \( K \in C^{2m}(\Omega \times \Omega) \) and \( p > \frac{n}{m} \). \( \rho \geq 1 \). Then,

\[
\mathcal{C}^2_K \leq D_{\Omega, p}^m(\sum_{i=N+1}^\infty \lambda_i^{2}(1-\theta)^{1/2}) \max_{|\alpha|=m} \|D^\alpha u\|_{L^p(\Omega)} \|D^\alpha f\|_{L^p(\Omega)}^\theta + D_{\Omega, p}^m \sum_{i=N+1}^\infty \lambda_i^{2}^{-1/2}
\]

where \( \theta = (1 + \frac{m}{n} - \frac{2}{p})^{-1} \) and

\[
D_{\Omega, p}^m = \sup_{u \in L^1(\Omega)} \frac{\|u\|_{L^1(\Omega)}}{\|u\|_{L^2(\Omega)}}
\]

is an optimal constant in the Gagliardo-Nirenberg inequality for the domain \( \Omega \times \Omega \).

For \( p = +\infty \) we have \( \mathcal{C}^2_K = O((\sum_{i=N+1}^\infty \lambda_i^{2})^{\frac{m}{n}}) \). For infinitely differentiable kernels, the latter implies \( \mathcal{C}^2_K = O((\sum_{i=N+1}^\infty \lambda_i^{2})^{1-\epsilon}) \) for any \( \epsilon > 0 \).

# Proof of the Main Theorem

Let \( \mathcal{H}_K \) be a reproducing kernel Hilbert space (RKHS) defined by \( K \). This space is a completion of the span of \( \{K(x, \cdot) \mid x \in \Omega\} \) with the inner product \( \langle K(x, \cdot), K(y, \cdot) \rangle_{\mathcal{H}_K} = K(x, y) \). Also, it can be characterized by the following proposition, which is equivalent to Theorem 4.12 from [3] and whose original version can be found in [2].

**Proposition 2.1 ([2, 3]).** Let \( \{\lambda_i\}_{i=1}^{\infty} \) be the set of all positive eigenvalues of \( \mathcal{O}_K \) (counting multiplicities) with corresponding orthogonal unit eigenvectors \( \{\phi_i\}_{i=1}^{\infty} \). Then, \( \mathcal{H}_K \) equals

\[
\mathcal{O}_K^{1/2}[L^2(\Omega)] = \left\{ \sum_{i=1}^{\infty} a_i \phi_i \mid \left[ \frac{a_i}{\sqrt{\lambda_i}} \right]_{i=1}^{\infty} \in l^2 \right\} \subseteq \mathcal{C}(\Omega)
\]
with the inner product \( \langle \sum_{i=1}^{\infty} a_i \varphi_i, \sum_{i=1}^{\infty} b_i \varphi_i \rangle_{\mathcal{H}_K} = \sum_{i=1}^{\infty} \frac{a_i b_i}{\| u \|_\alpha} \). For any \( f \in \mathcal{H}_K \),
\[
\| f \|_{L_\infty(\Omega)} \leq C_K \| f \|_{\mathcal{H}_K},
\]
where \( C_K = \max_{x, y \in \Omega} K(x, y) \).

We will use that proposition throughout our proof.

For any \( f \in C(\Omega) \), an internal point \( x \in \Omega \), \( h \in \mathbb{R}^n \) and \( \alpha \in (\mathbb{N} \cup \{0\})^n \), let us denote
\[
\delta^{(\alpha, \beta)}_h[f](x) = \sum_{\beta \leq \alpha} (-1)^{\beta} \binom{\alpha}{\beta} f(x_1 + \beta_1 h_1, \ldots, x_n + \beta_n h_n)
\]
where \( \binom{n}{\beta} \) denotes \( \beta_i \leq \alpha_i, i = 1, \ldots, n \). For a kernel \( K \in C(\Omega \times \Omega) \), we have
\[
\delta^{(\alpha, \beta)}_{(h, h')}[K](x, x) = \sum_{\beta, \beta' \leq \alpha, \beta' \leq \alpha} (-1)^{|\beta| + |\beta'|} \binom{\alpha}{\beta} \binom{\alpha}{\beta'} K(x_1 + \beta_1 h_1, \ldots, x_n + \beta_n h_n, x_1 + \beta'_1 h_1, \ldots, x_n + \beta'_n h_n)
\]
If \( \delta^{(\alpha, \beta)}_h \) exists, let us denote
\[
D_\alpha(x) = \delta^{(\alpha, 0)}_h K(x, y)_{y=x}.
\]
Note that \( \delta^{(\alpha, \beta)}_h \) is a finite difference operator of a higher order. Its well-known property is given below.

**Proposition 2.2.** If \( f \in C^{(\alpha)}(\Omega) \), then \( \delta^{(\alpha)}_h[f](x) = \partial^\alpha f(x) h^\alpha + r(x, h) \) where \( |r(x, h)| \leq C(x, h) \| h \|^{(\alpha)} \) and
\[
\lim_{h \to 0} C(x, h) = 0.
\]

For symmetric functions, \( \delta^{(\alpha, \alpha)}_{(h, h')} [F](x, x) \) satisfies a finer property.

**Lemma 2.3.** Let \( F \in C^{(\alpha)}(\Omega \times \Omega) \) satisfy \( F(x, y) = F(y, x) \). Then, for any \( \alpha \in (\mathbb{N} \cup \{0\})^n : |\alpha| = k \), we have
\[
\delta^{(\alpha, \alpha)}_{(h, h')} [F](x, x) = \partial^{(\alpha)}_x \partial^{(\alpha)}_y F|_{(x, x)} h^\alpha (h')^\alpha + r(x, h, h'),
\]
where
\[
|r(x, h, h')| \leq C_1(\alpha, \alpha) \| h \|^{(\alpha)} \| h' \|^{(\alpha)} + C_2(\alpha, \alpha) \| h \|^{(\alpha)} \| h' \|^{(\alpha)} + C(\alpha, \alpha) \| h \|^{(\alpha)} \| h' \|^{(\alpha)}
\]
and
\[
\lim_{(h, h') \to (0, 0)} C_1(\alpha, \alpha) = 0, \lim_{(h, h') \to (0, 0)} C_2(\alpha, \alpha) = 0, \lim_{(h, h') \to (0, 0)} C(\alpha, \alpha) = 0.
\]

**Proof.** A symbol \( f(h, h') = o(g(h, h')) \) denotes \( \lim_{(h, h') \to (0, 0)} \frac{f(h, h')}{g(h, h')} = 0. \) Let us denote
\[
q(h, h') = F(x + h, x + h') - \sum_{|\eta|, |\gamma|, |\eta'|, |\gamma'| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma - \sum_{|\eta|, |\gamma| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma - \sum_{|\eta|, |\gamma| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma
\]
\[
\sum_{|\eta|, |\gamma| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma
\]
\[
\sum_{|\eta|, |\gamma| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma
\]
\[
\sum_{|\eta|, |\gamma| \leq k} \frac{1}{\eta! \gamma!} \partial^{(\eta)}_x \partial^{(\gamma)}_y F|_{(x, x)} h^\eta h'^\gamma
\]
We will prove that \( q(h, h') = o(\|h'\|^k \|h\|^k) \). First, note that \( \partial_h^\alpha q(h, h') \), for \(|\alpha| \leq k\), reads as

\[
\partial_h^\alpha q(h, h') = \partial_h^\alpha F|_{(x+h,x+h')} - \sum_{q:|q| \leq k} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} - \sum_{q:|q| \leq k} \frac{(\partial_x^{q+\alpha} F|_{(x+h,x)} - \sum_{q:|q| \leq k} \frac{1}{\gamma!} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q)}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} - \sum_{\gamma:|\gamma| \leq k} \frac{(\partial_x^\gamma \partial_y^\gamma F|_{(x+h,x)} - \sum_{q:|q| \leq k} \frac{1}{\gamma!} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q)}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!}
\]

and therefore,

\[
\partial_h^\alpha q(0, h') = \partial_h^\alpha F|_{(x,x+h')} - \sum_{\gamma:|\gamma| \leq k} \frac{\partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} - \sum_{\gamma:|\gamma| \leq k} \frac{(\partial_x^\gamma \partial_y^\gamma F|_{(x,x+h)} - \sum_{q:|q| \leq k} \frac{1}{\gamma!} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q)}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} = 0
\]

Using \( q(\cdot, h') \in C^k(\Omega) \) and Taylor’s expansion around \( h = 0 \), we obtain

\[
q(h, h') = \sum_{|\alpha| = k} \partial_h^\alpha q(\chi h, h') \frac{h^\alpha}{\alpha!}
\]

where \( \chi \in (0, 1) \).

For \(|\alpha| = k\), we have

\[
\partial_h^\alpha q(\chi h, h') = \partial_h^\alpha F|_{(x+\chi h,x+h')} - \sum_{\gamma:|\gamma| \leq k} \frac{\partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} - \sum_{\gamma:|\gamma| \leq k} \frac{(\partial_x^\gamma \partial_y^\gamma F|_{(x+\chi h,x)} - \sum_{q:|q| \leq k} \frac{1}{\gamma!} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x)} h^q)}{\gamma!} \partial_y^q F|_{(x,x)} \frac{h^q h'^y}{\eta!} - \sum_{\gamma:|\gamma| \leq k} \frac{(\partial_x^\gamma \partial_y^\gamma F|_{(x+\chi h,x+h')} - \sum_{q:|q| \leq k} \frac{1}{\gamma!} \partial_x^{q+\alpha} \partial_y^q F|_{(x,x+h')} h^q)}{\gamma!} \partial_y^q F|_{(x,x+h')} \frac{h^q h'^y}{\eta!}
\]

If we denote \( R(h') = \partial_h^\alpha F|_{(x+\chi h,x+h')} - \partial_h^\alpha F|_{(x,x+h')} \), then, by Taylor’s expansion theorem, we have \( R(h') - \sum_{\gamma:|\gamma| \leq k} \frac{\partial_x^\gamma R(0) h'^y}{\gamma!} = o(\|h'\|^k) \). The latter expression for \( \partial_h^\alpha q(\chi h, h') \) exactly equals \( R(h') - \sum_{\gamma:|\gamma| \leq k} \frac{\partial_x^\gamma R(0) h'^y}{\gamma!} \) and we conclude

\[
q(h, h') = \sum_{|\alpha| = k} \partial_h^\alpha q(\chi h, h') \frac{h^\alpha}{\alpha!} = o(\|h\|^k \|h'\|^k).
\]
Thus, we proved that

\[
F(x + h, x + h') = \sum_{\eta ; |\eta| \leq k} A_{\eta, y} h^\eta (h')^\eta + \sum_{y ; |y| \leq k} a_{y}(h) h^y + \sum_{\eta ; |\eta| \leq k} a_{\eta}(h') h^\eta + q(h, h'),
\]

where \( A_{\eta, y} = \frac{1}{\eta y!} \frac{\partial^2 \partial y F(y, x)}{x^0 \partial y} \) and \( a_{y}(h) = \frac{\partial^2 \partial y F(y, x)}{x^0 \partial y} \). After plugging in the expression (4) into (3), we have \((\odot \text{ denotes the Hadamard product})\)

\[
\delta^{(a, a)} [(x, x)] = \sum_{\beta, \beta' : \beta, \beta' \leq a, \beta' \leq h} (-1)^{|\beta| + |\beta'|} \left( \frac{\alpha}{\beta} \right) \left( \frac{\alpha'}{\beta'} \right) a_{y}(y) h^y + o(||h||^k ||h'||^k).
\]

Note that

\[
\sum_{\beta, \beta' : \beta, \beta' \leq a, \beta' \leq h} (-1)^{|\beta| + |\beta'|} \left( \frac{\alpha}{\beta} \right) \left( \frac{\alpha'}{\beta'} \right) h^\eta = \frac{1}{\eta!} \frac{\partial^2 \partial y F(y, x)}{x^0 \partial y} h^\eta + o(||h||^k ||h'||^k).
\]

The expression that is in the RHS is just a finite difference of order \( a_i \) of \( f(x) = x^\eta \) for \( h = 1 \), due to \( \delta^{(a_i)} [x^\eta](0) = \sum_{\beta_i = 0}^{a_i} (1 - \beta_i) \beta_i^\eta \cdot \beta^{\eta'}(\beta')^\eta' \). It is well-known that \( \delta^{(a_i)} [x^\eta](x) = 0 \), if \( \eta_i < a_i \) and \( \delta^{(a_i)} [x^\eta](x) = \eta_i! \), if \( \eta_i = a_i \). Thus, we have

\[
\sum_{\beta, \beta' : \beta, \beta' \leq a, \beta' \leq h} (-1)^{|\beta| + |\beta'|} \left( \frac{\alpha}{\beta} \right) h^\eta = \frac{1}{\eta!} \frac{\partial^2 \partial y F(y, x)}{x^0 \partial y} h^\eta + o(||h||^k ||h'||^k).
\]
Therefore,
\[
\delta((a|\infty)_{(h_1, \ldots, h_n)}[F](x, x) = \frac{(\hat{a})^2}{(\hat{a})^2} \delta_{(h_1, \ldots, h_n)} K(x, y) \bigr|_{y=x} h^a (h^a) + o(|h^n| h^a + o(|h^n| h^a) + o(|h^n| h^a)
\]
From the latter, the statement of Lemma directly follows. \]

The following lemma is a direct consequence of Theorem 1 from [16]. We give here its proof for the sake of completeness.

**Lemma 2.4.** Let \( K \in C^2(\Omega \times \Omega) \) and \( x \in \Omega \) be fixed. Let \( \lambda_{i} \) be a multiset of all positive eigenvalues of \( \Omega_K \) (counting multiplicities). Then, \( \partial_{\varphi}^2 K(x, \cdot) \in \mathcal{H}_K \) and \( \|\partial_{\varphi}^2 K(x, \cdot)\|^2_{\mathcal{H}_K} = D_\alpha(x) = \sum_{i=1}^\infty \lambda_i (\partial_{\varphi}^2 \phi_i(x))^2. \)

**Proof.** Let us choose some sequence \( \{h_i\}_{i=1}^\infty \) such that \( \lim_{i \to \infty} h_i = 0 \) and let
\[
f_i(y) = \delta((a|\infty)_{(h_1, \ldots, h_i)} [K(z, y)](x) \in \mathcal{H}_K
\]
where the finite difference operator \( \delta_h \) is applied onto the first argument. The inner product between \( f_i \) and \( f_j \) equals:
\[
\langle f_i, f_j \rangle_{\mathcal{H}_K} = \frac{\delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, x)}{h_i^{a|\infty} h_j^{a|\infty}}.
\]
Therefore,
\[
\|f_i - f_j\|^2_{\mathcal{H}_K} = \frac{\delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, x)}{h_i^{a|\infty} h_j^{a|\infty}} + \frac{\delta((a|\infty)_{(h_1, \ldots, h_j)} [K](x, x)}{h_j^{a|\infty} h_j^{a|\infty}} - 2 \frac{\delta((a|\infty)_{(h_1, \ldots, h_i, h_j)} [K](x, x)}{h_i^{a|\infty} h_j^{a|\infty}}.
\]
From Lemma 2.3 we obtain that for any \( \varepsilon > 0 \) there exists \( N_\varepsilon > 0 \) such that \( \frac{\delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, x)}{h_i^{a|\infty}} - D_\alpha(x) < \varepsilon \) and \( \frac{\delta((a|\infty)_{(h_1, \ldots, h_i, h_j)} [K](x, x)}{h_j^{a|\infty}} - D_\alpha(x) < \varepsilon \) whenever \( i > N_\varepsilon, j > N_\varepsilon \). Therefore, \( \|f_i - f_j\|^2_{\mathcal{H}_K} \leq 4 \varepsilon \) if \( i > N_\varepsilon, j > N_\varepsilon \). The latter means that \( \{f_i\} \subseteq \mathcal{H}_K \) is a Cauchy sequence. From the completeness of \( \mathcal{H}_K \) we conclude that \( f_i \to \mathcal{H}_K f \) where \( f \in \mathcal{H}_K \). From Proposition 2.1 we conclude that \( f_i \) uniformly converges to \( f \). By construction, the pointwise limit of \( f_i = \delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, y)](x) \) is \( \partial_{\varphi}^2 K(x, \cdot) \). Therefore, \( f_i \to \mathcal{H}_K \partial_{\varphi}^2 K(x, \cdot) \) and \( \partial_{\varphi}^2 K(x, \cdot) \in \mathcal{H}_K \).

Let \( f_\varepsilon(y) = \delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, y)](x) \) for \( h = (h_1, \ldots, h) \). In fact, we have just proved that \( \lim_{h \to 0} \delta((a|\infty)_{(h_1, \ldots, h_i)} [K](x, y)](x) = \partial_{\varphi}^2 K(x, \cdot) \) in \( \mathcal{H}_K \). According to Mercer’s theorem, we have
\[
\lim_{N \to \infty} \sup_{z, y \in \Omega} |K(z, y) - \sum_{i=1}^N \lambda_i \phi_i(z) \phi_i(y)| = 0.
\]
A sum of $k$ uniformly convergent function series equals a uniformly convergent series of the corresponding $k$-sums, i.e.
\[
\frac{f_k(y)}{h_{|x|}} = \frac{\delta_{|x|}^\infty}{h_{|x|}} \left[ \frac{\lim_{N \to \infty} \sum_{i=1}^N \lambda_i \phi_i(z) \phi_i(y)}{h_{|x|}} \right](x) = \lim_{N \to \infty} \sum_{i=1}^N \lambda_i \left( \frac{\delta_{|x|}^\infty}{h_{|x|}} \phi_i \right)(y)
\]

Therefore, \( \frac{f_k(y)}{h_{|x|}} = \sum_{i=1}^\infty \lambda_i \left( \frac{\delta_{|x|}^\infty}{h_{|x|}} \phi_i \right)(y) \) and the latter convergence is uniform over $y$.

Therefore, \( \int \frac{f_k(y)}{h_{|x|}} \phi_i(y) d\mu(y) = \lambda_i \left( \frac{\delta_{|x|}^\infty}{h_{|x|}} \phi_i \right)(x) \). A uniform convergence of \( \frac{f_k(y)}{h_{|x|}} \) to \( \delta_x^\infty K(x, y) \) as $h \to 0$ implies
\[
\lambda_i \partial_x^\alpha \phi_i(x) = \lim_{h \to 0} \frac{\delta_{|x|}^\infty [\phi_i(x)]}{h_{|x|}} = \int \delta_x^\infty K(x, y) \phi_i(y) d\mu(y).
\]

Since \( \delta_x^\infty K(x, \cdot) \in \mathcal{H}_K \), using Proposition 2.1, we conclude:
\[
\| \delta_x^\infty K(x, \cdot) \|_{\mathcal{H}_K}^2 = \sum_{i=1}^\infty \lambda_i^2 \left( \frac{\partial_x^\alpha \delta_x^\infty \phi_i(x)}{\lambda_i} \right)^2
\]

Since \( \lim_{h \to 0} \left( \frac{f_k}{h_{|x|}}, \frac{f}{h_{|x|}} \right)_{\mathcal{H}_K} = \lim_{h \to 0} \frac{\Delta^\infty_{|x|}(K)(x)}{h_{|x|}} = D_\alpha(x) \) we finally obtain
\[
D_\alpha(x) = \| \partial_x^\infty K(x, \cdot) \|_{\mathcal{H}_K}^2 = \sum_{i=1}^\infty \lambda_i (\partial_x^\alpha \phi_i(x))^2
\]

**Lemma 2.5.** Let \( K(x, y) \in C^{2m}(\Omega \times \Omega) \) and \( \{\lambda_i\} \) be a multiset of all positive eigenvalues of \( O_K \) (counting multiplicities). Then,
\[
\partial_x^\alpha \partial_y^\beta K(x, y) = \sum_{i=1}^\infty \lambda_i \partial_x^\alpha \phi_i(x) \partial_y^\beta \phi_i(y)
\]

for $|\alpha| \leq m$ and $|\beta| \leq m$.

**Proof.** Again, since \( \lambda_i \phi_i(x) = \int \Omega K(x, y) \phi_i(y) d\mu \), we conclude \( \lambda_i \partial_x^\alpha \phi_i(x) = \int \Omega \partial_x^\alpha K(x, y) \phi_i(y) d\mu \in C(\Omega) \) for $|\alpha| \leq m$. From Lemma 2.4 and Dini’s theorem we conclude that the series
\[
\sum_{i=1}^\infty \lambda_i |\partial_x^\alpha \phi_i(x) | \partial_y^\beta \phi_i(y)| \leq \frac{1}{2} \sum_{i=1}^\infty \lambda_i (\partial_x^\alpha \phi_i(x))^2 + \lambda_i (\partial_y^\beta \phi_i(y))^2
\]

is absolutely and uniformly convergent. Therefore, we can differentiate the function series, and conclude
\[
\sum_{i=1}^\infty \lambda_i \partial_x^\alpha \phi_i(x) \partial_y^\beta \phi_i(y) = \partial_x^\alpha \partial_y^\beta \left( \sum_{i=1}^\infty \lambda_i \phi_i(x) \phi_i(y) \right) = \partial_x^\alpha \partial_y^\beta K(x, y).
\]

Let us denote
\[
K_N(x, y) = K(x, y) - \sum_{i=1}^N \lambda_i \phi_i(x) \phi_i(y)
\]

and
\[
K_N^\alpha\beta(x, y) = \partial_x^\alpha \partial_y^\beta K(x, y) - \sum_{i=1}^N \lambda_i \partial_x^\alpha \phi_i(x) \partial_y^\beta \phi_i(y)
\]
Lemma 2.6. Let $K(x, y) \in C^{2m}(\Omega \times \Omega)$ for compact $\Omega \subseteq \mathbb{R}^n$ and $\{\lambda_i\}$ be a multiset of all positive eigenvalues of $O_K$ (counting multiplicities). Then, for any $|\alpha| \leq m$, $|\beta| \leq m$, we have

$$|K_N^{\alpha,\beta}(x, y)| \leq K_N^{\alpha,\alpha}(x, x)^{1/2} K_N^{\beta,\beta}(y, y)^{1/2} \leq D_\alpha(x)^{1/2} D_\beta(y)^{1/2}.$$ 

Proof. From Lemmas 2.4 and 2.5 we have

$$\partial_x^\alpha K_N(x, \cdot) = \sum_{i=N+1}^\infty \lambda_i \partial_x^\alpha \phi_i(x) \phi_i(y) \in \mathcal{H}_K.$$ 

Using Proposition 2.1, and again, Lemma 2.5, we obtain

$$\partial_x^\alpha \partial_y^\beta K_N(x, y) = \sum_{i=N+1}^\infty \lambda_i \partial_x^\alpha \phi_i(x) \partial_y^\beta \phi_i(y) = \langle \partial_x^\alpha K_N(x, \cdot), \partial_y^\beta K_N(y, \cdot) \rangle_{\mathcal{H}_K}.$$ 

Finally, the Cauchy-Schwartz inequality gives us

$$|\langle \partial_x^\alpha K_N(x, \cdot), \partial_y^\beta K_N(y, \cdot) \rangle_{\mathcal{H}_K}| \leq \|\partial_x^\alpha K_N(x, \cdot)\|_{\mathcal{H}_K} \cdot \|\partial_y^\beta K_N(y, \cdot)\|_{\mathcal{H}_K} = K_N^{\alpha,\alpha}(x, x)^{1/2} K_N^{\beta,\beta}(y, y)^{1/2}.$$ 

Note that

$$K_N^{\alpha,\alpha}(x, y) = \sum_{i=N+1}^\infty \lambda_i \partial_x^\alpha \phi_i(x)^2 \leq \sum_{i=1}^\infty \lambda_i \partial_x^\alpha \phi_i(x)^2 = D_\alpha(x).$$

Therefore, we have

$$|K_N^{\alpha,\beta}(x, y)| \leq D_\alpha(x)^{1/2} D_\beta(y)^{1/2}.$$ 

\[\square\]

Proof of Theorem 1.1. The tightness of our bounds strongly depends on the constant $C_{\Omega, \rho}$ in the Gagliardo-Nirenberg inequality, which reads as [1, 9]

$$\|u\|_{L^\infty(\Omega)} \leq C_{\Omega, \rho} \|u\|_{L^1(\Omega)}^{1-\theta} \cdot \|D^m u\|_{L^\rho(\Omega)}^\theta + C_{\Omega, \rho} \|u\|_{L^1(\Omega)},$$

where $\theta(\frac{\rho}{n} - m) + (1 - \theta)n = 0$ and $\|D^m u\|_{L^\rho(\Omega)} = \max_{|\alpha| \leq m} \|\partial_\alpha u(x)\|_{L^\rho(\Omega)}$. Thus, $\theta = \frac{\rho}{n-m \rho + m} = (1 + \frac{m}{n} - \frac{1}{\rho})^{-1}$.

Using $\sup |K_N(x, y)| \leq \sup K_N(x, x)^{1/2} K_N(y, y)^{1/2} = \sup K_N(x, x)$ and the Gagliardo-Nirenberg inequality we have

$$C_{\Omega, \rho} = \|K_N(x, x)\|_{L^\infty(\Omega \times \Omega)} = \|K_N(x, x)\|_{L^\infty(\Omega)} \leq C_{\Omega, \rho} \|K_N(x, x)\|_{L^1(\Omega)}^{1-\theta} \cdot \|D^m K_N(x, x)\|_{L^\rho(\Omega)}^\theta + C_{\Omega, \rho} \|K_N(x, x)\|_{L^1(\Omega)} \leq C_{\Omega, \rho} \left( \sum_{i=N+1}^{\infty} \lambda_i \right)^{1-\theta} \cdot \|D^m K_N(x, x)\|_{L^\rho(\Omega)}^\theta + C_{\Omega, \rho} \sum_{i=N+1}^{\infty} \lambda_i.$$ 

Lemma 2.6 gives us

$$|\partial_x^\alpha \partial_y^\beta [K_N(x, x)]| = \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} k_{N}^{\beta,\alpha-\beta}(x, x) \leq \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} D_\beta(x)^{1/2} D_{\alpha-\beta}(x)^{1/2}. $$
Therefore, we have
\[
C_K^N \leq C_{\Omega,p} \left( \sum_{i=N+1}^{\infty} \lambda_i \right)^{1-\theta} \cdot \max_{\alpha,|\alpha|=m} \left( \sum_{\beta \leq \alpha} \left( \frac{a}{\beta} \right) \right) \|D_{\beta}D_{\alpha-\beta}\|_{L_p(\Omega)}^{\theta} + C_{\Omega,p} \sum_{i=N+1}^{\infty} \lambda_i.
\]

Theorem proved.  \( \square \)

**Proof of Theorem 1.2.** Another version of the Gagliardo-Nirenberg inequality, now for the domain \( \Omega \times \Omega \), is
\[
\|u\|_{L^\infty(\Omega \times \Omega)} \leq D_{\Omega,p}\|u\|_{L^2(\Omega \times \Omega)}^{1-\theta} \cdot \|D^m u\|_{L^\infty(\Omega \times \Omega)}^\theta + D_{\Omega,p}\|u\|_{L^2(\Omega \times \Omega)},
\]
where \( \theta(\frac{n}{p} - m) + (1 - \theta) \frac{n}{2} = 0 \). Therefore, \( \theta = (1 + \frac{2m}{n} - \frac{2}{p})^{-1} \). For \( u(x,y) = K_N(x,y) \), we have
\[
C_K^N = \|K_N(x,y)\|_{L^\infty(\Omega \times \Omega)} \leq D_{\Omega,p}\|K_N\|_{L^2(\Omega \times \Omega)}^{1-\theta} \cdot \|D^m K_N\|_{L^\infty(\Omega \times \Omega)}^\theta + D_{\Omega,p}\|K_N\|_{L^2(\Omega \times \Omega)}.
\]
Using Lemma 2.6 we obtain
\[
\|D^m K_N\|_{L^p} = \max_{|\alpha|+|\beta|=m} \|\partial_x^\alpha \partial_y^\beta [K_N(x,y)]\|_{L^p} \leq \max_{|\alpha|+|\beta|=m} \|\sqrt{D_{\alpha}}\|_{L^p} \|\sqrt{D_{\beta}}\|_{L^p}.
\]
Therefore,
\[
C_K^N \leq D_{\Omega,p} \left( \sum_{i=N+1}^{\infty} \lambda_i^{2(1-\theta)/2} \cdot \max_{|\alpha|+|\beta|=m} \|\sqrt{D_{\alpha}}\|_{L^p} \|\sqrt{D_{\beta}}\|_{L^p} + D_{\Omega,p} \left( \sum_{i=N+1}^{\infty} \lambda_i^2 \right)^{1/2}.
\]
Theorem proved.  \( \square \)

3 APPLICATIONS

**Bounding the kernel of \( O_K^\gamma \).** For \( \gamma > 0 \), let us denote
\[
K^\gamma(x,y) = \sum_{i=1}^{\infty} \lambda_i^\gamma \phi_i(x)\phi_i(y).
\]
In general, checking the condition \( \sup_{x \in \Omega} K^\gamma(x,x) < \infty \) requires the study of eigenvectors \( \phi_i \). For kernels that appear in applications [3, 5], a concrete form of eigenvectors is known only in few cases [4]. In the current paper we are interested in information that can be extracted from a behavior of eigenvalues \( \{\lambda_i\} \). Let us formulate one example of such a sufficient condition.

Note that if \( \sum_{i=1}^{\infty} \lambda_i^2 y < \infty \), then \( \sum_{i=1}^{\infty} \lambda_i^\gamma \phi_i(x)\phi_i(y) \in L^2(\Omega \times \Omega) \). In a special case \( \gamma = \frac{1}{2} \) we have \( \sum_{i=1}^{\infty} \lambda_i^{y} = \text{Tr}(O_K) < \infty \). Therefore, \( K^\gamma \in L^2(\Omega \times \Omega) \) for \( \gamma \in [\frac{1}{2}, 1] \). The boundedness of \( K^\gamma \) on the diagonal, i.e. \( \sup_{x \in \Omega} K^\gamma(x,x) < \infty \) is equivalent to \( K^\gamma \in C(\Omega \times \Omega) \). Indeed, if \( K^\gamma(x,x) < C \), then \( f_N(x) = \sum_{i=1}^{\infty} \lambda_i^\gamma \phi_i(x)^2 \) is a monotonically increasing sequence of nonnegative continuous functions on a compact set \( \Omega \), bounded by \( C \). Then, by monotone convergence theorem, \( \{f_N\} \) uniformly converges to a continuous function \( K^\gamma(x,x) \). From the uniform convergence of the series \( \sum_{i=1}^{\infty} \lambda_i^\gamma \phi_i(x)^2 \) it is straightforward that \( \sum_{i=1}^{\infty} \lambda_i^\gamma \phi_i(x)^2 < \frac{1}{2} \sum_{i=1}^{\infty} \lambda_i^\gamma (\phi_i(x)^2 + \phi_i(y)^2) \) is also uniformly convergent to a continuous function.
Theorem 3.1. Let $K \in \mathcal{C}^{2m}(\Omega \times \Omega)$ and $\gamma \in (0, 1)$. Then, for $\sup_{x \in \Omega} K^{1-\gamma}(x, x) < \infty$ it is sufficient to have

$$\sum_{i=N+1}^{\infty} \lambda_i^2 = o\left(\frac{m_{\gamma}}{m}\right)$$

and

$$\sum_{N=1}^{\infty} \left( \sum_{i=N+1}^{\infty} \lambda_i^2 \right)^{\frac{m}{m_{\gamma}}} \left( \lambda_{N+1}^{-\gamma} - \lambda_N^{-\gamma} \right) < \infty.$$

**Proof.** Let us denote $K_N(x) = \sum_{i=N}^{\infty} \lambda_i \phi_i(x)^2$. Then

$$K^{1-\gamma}(x, x) = \sum_{i=1}^{\infty} \lambda_i^{-\gamma} \lambda_i \phi_i(x)^2 = \sum_{i=1}^{\infty} \lambda_i^{-\gamma} (K_i(x) - K_{i+1}(x)) =$$

using summation by parts formula

$$= \lambda_1^{-\gamma} K_1(x) - \lim_{N \to +\infty} \lambda_N^{-\gamma} K_{N+1}(x) + \sum_{i=2}^{\infty} K_i(x) (\lambda_i^{-\gamma} - \lambda_{i-1}^{-\gamma}) \leq$$

$$\lambda_1^{-\gamma} D_K^2 + \lim_{N \to +\infty} \lambda_N^{-\gamma} (C_N^K)^2 + \sum_{N=1}^{\infty} (C_N^K)^2 (\lambda_{N+1}^{-\gamma} - \lambda_N^{-\gamma})$$

In Theorem 1.2 it was shown that for $K \in \mathcal{C}^{2m}(\Omega \times \Omega)$ we have $(C_N^K)^2 \leq C \left(\sum_{i=N+1}^{\infty} \lambda_i^2\right)^{\frac{m}{m_{\gamma}}}$. Therefore,

$$\lim_{N \to +\infty} \lambda_N^{-\gamma} (\sum_{i=N+1}^{\infty} \lambda_i^2)^{\frac{m}{m_{\gamma}}} = 0$$

and

$$\sum_{N=1}^{\infty} \left( \sum_{i=N+1}^{\infty} \lambda_i^2 \right)^{\frac{m}{m_{\gamma}}} \left( \lambda_{N+1}^{-\gamma} - \lambda_N^{-\gamma} \right) < \infty$$

is sufficient for $\sup_{x \in \Omega} K^{1-\gamma}(x, x) < \infty$. \hfill \qed

Let us show how to apply the latter bound for infinitely differentiable kernels. In the case of an infinitely differentiable kernel, we have

$$(C_N^K)^2 \leq C \sum_{i=N+1}^{\infty} \lambda_i^2 \lambda_i^{-\gamma}$$

for any $\gamma > 0$. Let us additionally assume that eigenvalues of $O_K$ are rapidly vanishing, i.e. $\sum_{i=N+1}^{\infty} \lambda_i^2 = O(\lambda_{N+1}^2)$ and $\sum_{i=1}^{\infty} \lambda_i^2 < \infty$ for any $\epsilon > 0$. Note that these conditions are satisfied for the Gaussian kernel on a box or a ball in $\mathbb{R}^n$, analytic kernels on a finite interval [7]. Let $\gamma \in (0, 1)$. We have $\lambda_N^{-\gamma} (C_N^K)^2 \leq C \lambda_N^{-\gamma} (\sum_{i=N+1}^{\infty} \lambda_i^2)^{\frac{m}{m_{\gamma}}} = O(\lambda_{N+1}^2 \lambda_N^{-\gamma}) \to 0$, since $\gamma$ can be chosen to satisfy $\gamma < 1-2\epsilon$. Also, $\sum_{i=N+1}^{\infty} (C_N^K)^2 (\lambda_{N+1}^{-\gamma} - \lambda_N^{-\gamma}) \leq C \sum_{N=1}^{\infty} (\sum_{i=N+1}^{\infty} \lambda_i^2)^{\frac{m}{m_{\gamma}}} \lambda_{N+1}^{-\gamma} \leq C \sum_{N=1}^{\infty} \lambda_{N+1}^{-\gamma} \lambda_{N+1}^{-\gamma} < \infty$. Thus, for $\gamma \in (0, 1)$, $K^\gamma$ is bounded and continuous.

**Bounding the supremum norm of eigenvectors.** The condition

$$\sup_N \| \phi_N \|_{L^\infty(\Omega)} < \infty$$

is popular in various statements concerning Mercer kernels, though it is believed to be hard to check. Discussions of that issue can be found in [8, 12, 15].

Since $\lambda_{N+1} \phi_{N+1}(x)^2 \leq K(x, x) - \sum_{i=1}^{N} \lambda_i \phi_i(x)^2$, we conclude

$$\| \phi_{N+1} \|_{L^\infty(\Omega)} \leq \lambda_{N+1}^{-1/2} \sqrt{C_N^K}.$$
Thus, any upper bound for $C_N^N$ leads to an upper bound of $\|\phi_{N+1}\|_{L^\infty(\Omega)}$. For a uniform boundedness of $\|\phi_{N+1}\|_{L^\infty(\Omega)}$ we need $C_N^N = O(\lambda_{N+1})$. Unfortunately, RHS of our bounds are not $O(\lambda_{N+1})$, though they can be used to show a moderate growth rate of $\|\phi_{N+1}\|_{L^\infty(\Omega)}$.

REFERENCES

[1] Brezis, H. Function Analysis, Sobolev Spaces and Partial Differential Equations. 01 2010.
[2] Cucker, F., and Smale, S. On the mathematical foundations of learning. Bulletin of the American Mathematical Society 39 (2001), 1–49.
[3] Cucker, F., and Zhou, D.X. Learning Theory: An Approximation Theory Viewpoint. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2007.
[4] De Vito, E., Mücke, N., and Rosasco, L. Reproducing kernel hilbert spaces on manifolds: Sobolev and diffusion spaces. Analysis and Applications 19 (08 2020).
[5] Guo, X., Li, L., and Wu, Q. Modeling interactive components by coordinate kernel polynomial models. Mathematical Foundations of Computing 3, 4 (2020), 263–277.
[6] König, H. Eigenvalue Distribution of Compact Operators. Eigenvalue Distribution of Compact Operators. Birkhäuser Basel, 1986.
[7] Little, G., and Reade, J. B. Eigenvalues of analytic kernels. SIAM Journal on Mathematical Analysis 15, 1 (1984), 133–136.
[8] Mendelson, S., and Neeman, J. Regularization in kernel learning. The Annals of Statistics 38, 1 (2010), 526 – 565.
[9] Nieberg, L. An extended interpolation inequality. Annali della Scuola Normale Superiore di Pisa - Classe di Scienze Ser. 3, 20, 4 (1966), 733–737.
[10] Schölkopf, B., Smola, A., and Müller, K.-R. Kernel principal component analysis. In Artificial Neural Networks — ICANN’97 (Berlin, Heidelberg, 1997), W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, Eds., Springer Berlin Heidelberg, pp. 583–588.
[11] Steinwart, I., and Christmann, A. Support Vector Machines. Information Science and Statistics. Springer New York, 2008.
[12] Steinwart, I., and Download C. Mercer’s theorem on general domains: On the interaction between measures, kernels, and rkhs. Constructive Approximation 35, 3 (Jun 2012), 365–417.
[13] Takhanov, R. Dimension reduction as an optimization problem over a set of generalized functions, 2019.
[14] Wahba, G. Spline Models for Observational Data. Society for Industrial and Applied Mathematics, 1990.
[15] Zhou, D.-X. The covering number in learning theory. J. Complex. 18, 3 (Sept. 2002), 739–767.
[16] Zhou, D.-X. Derivative reproducing properties for kernel methods in learning theory. Journal of Computational and Applied Mathematics 220, 1 (2008), 456–463.