JOURNAL OF COMPLEXITY **18**, 739–767 (2002)
doi:10.1006/jcom.2002.0635

# The Covering Number in Learning Theory[1]

Ding-Xuan Zhou

*Department of Mathematics, City University of Hong Kong,*
*Tat Chee Avenue, Kowloon, Hong Kong, China*
E-mail: mazhou@math.cityu.edu.hk

Received May 25, 2001; revised October 22, 2001; accepted October 26, 2001;
published online July 2, 2002

The covering number of a ball of a reproducing kernel Hilbert space as a subset of the continuous function space plays an important role in Learning Theory. We give estimates for this covering number by means of the regularity of the Mercer kernel $K$. For convolution type kernels $K(x,t) = k(x-t)$ on $[0,1]^n$, we provide estimates depending on the decay of $\hat{k}$, the Fourier transform of $k$. In particular, when $\hat{k}$ decays exponentially, our estimate for this covering number is better than all the previous results and covers many important Mercer kernels. A counter example is presented to show that the eigenfunctions of the Hilbert–Schmidt operator $L_K$ associated with a Mercer kernel $K$ may not be uniformly bounded. Hence some previous methods used for estimating the covering number in Learning Theory are not valid. We also provide an example of a Mercer kernel to show that $L_K^{1/2}$ may not be generated by a Mercer kernel. © 2002 Elsevier Science (USA)

*Key Words:* Learning Theory; covering number; Mercer kernel; reproducing kernel Hilbert space.

## 1. INTRODUCTION

Learning Theory studies learning objects from random samples. To estimate the (probability) error or the number of samples required for given confidence and error bound, **covering numbers** or entropy numbers are deeply involved and play an essential role (see [2, 4, 12]). For a compact set $S$ in a metric space and $\eta > 0$, the covering number $\mathcal{N}(S, \eta)$ is defined to be the minimal integer $m \in \mathbb{N}$ such that there exist $m$ disks with radius $\eta$ covering $S$.

In estimating the error and the number of required samples for kernel machine learning [2, 4, 6, 11], a *reproducing kernel Hilbert space* (RKHS) is often used and the covering number of a ball in such a space is needed to derive bounds.

[1]The author is supported by CERG Grant 9040649 and City University Grant 7001214.



0885-064X/02 $35.00
© 2002 Elsevier Science (USA)
All rights reserved.

Let $(X, d(\cdot,\cdot))$ be a compact metric space, and $v$ be a Borel measure on $X$. Denote the space of real square integrable functions on $X$ by $L^2(X)$, and the space of continuous functions by $C(X)$.

Let $K : X \times X \to \mathbb{R}$ be continuous, symmetric and positive definite, *i.e.*, for any finite set $\{x_1, \ldots, x_m\} \subset X$, the matrix $(K(x_i, x_j))_{i,j=1}^m$ is positive definite. We call $K$ a **Mercer kernel**. Define a Hilbert–Schmidt integral operator by means of this kernel as

$$L_K f(x) = \int_X K(x,t) f(t) dv(t), \qquad x \in X, \ \ f \in L^2(X). \tag{1.1}$$

Then $L_K$ is a positive, compact operator and its range lies in $C(X)$.

Let $\{\lambda_j\}_{j=1}^\infty$ denote the nonincreasing sequence of eigenvalues of $L_K$ and $\{\phi_j\}$ be corresponding eigenfunctions. Then

$$K(x,t) = \sum_{j=1}^\infty \lambda_j \phi_j(x) \phi_j(t), \tag{1.2}$$

where the series converges uniformly and absolutely.

The reproducing kernel Hilbert space $\mathscr{H}_K$ associated with the kernel $K$ is defined (see [1]) to be the closure of the linear span of the set of functions $\{K_x := K(x,\cdot) \colon x \in X\}$ with the inner product satisfying

$$\langle K_x, f \rangle_{\mathscr{H}_K} = f(x), \qquad \forall x \in X, \ \ f \in \mathscr{H}_K. \tag{1.3}$$

An equivalent definition can be given by means of the square root of $L_K$. Take $L_K^{1/2}$ to be the linear operator on $L^2(X)$ satisfying $L_K^{1/2} L_K^{1/2} = L_K$, *i.e.*, $L_K^{1/2}(\phi_j) = \sqrt{\lambda_j}\phi_j$ for each $j$. Then $\mathscr{H}_K = L_K^{1/2}(L^2(X))$, and $\|f\|_K = \|(L_K^{1/2})^{-1} f\|_{L^2(X)}$. This space can be imbedded into $C(X)$, and we denote the inclusion as $I_K : H_K \to C(X)$. For these facts, see [2].

Let $R > 0$ and $B_R$ be the ball of $\mathscr{H}_K$ with radius $R$:

$$B_R := \{f \in H_K \colon \|f\|_K \leqslant R\}.$$

Then $I_K(B_R)$ is a subset of $C(X)$. Denote its closure in $C(X)$ as $\overline{I_K(B_R)}$ which is a compact subset. We are interested in the covering number $\mathscr{N}(\overline{I_K(B_R)}, \eta)$, where $\eta > 0$.

What is also often used in learning theory is the covering numbers of discrete versions of $\overline{I_K(B_R)}$. Let $\mathbf{x} := \{x_1, \ldots, x_l\} \subset X$. Consider the restriction of functions in $\overline{I_K(B_R)}$ to these points:

$$\overline{I_K(B_R)}\,|\mathbf{x} := \{(f(x_i))_{i=1}^l \in \mathbb{R}^l \colon \ f \in \overline{I_K(B_R)}\}.$$

This is a bounded subset of $\mathbb{R}^l$. We denote $\mathscr{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta)$ as the covering number of this subset of $\mathbb{R}^l$ with the $\ell_\infty$ metric: $d((a_i)_{i=1}^l, (b_i)_{i=1}^l) := \max_{1 \leqslant i \leqslant l} |a_i - b_i|$.

It is easily seen that if $\overline{I_K(B_R)}$ is covered by the uninion of balls in $C(X)$ with radius $\eta$ and centers $\{f_j(x) \in C(X): j = 1, \ldots, m\}$, then $\overline{I_K(B_R)}|_{\mathbf{x}}$ is covered by the uninion of balls in $\mathbb{R}^l$ with radius $\eta$ and centers $\{(f_j(x_i))_{i=1}^l \in \mathbb{R}^l: j = 1, \ldots, m\}$. Therefore,

$$\mathcal{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta) \leqslant \mathcal{N}(\overline{I_K(B_R)}, \eta), \qquad \forall \eta > 0.$$

The other direction was studied by Pontil [7]. Suppose $X$ is a compact subset of $\mathbb{R}^n$ and $K \in C^2(X \times X)$. Then it was shown in [7] that

$$\mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \mathcal{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta - R\|K\|_{C^2(X \times X)}^{1/2} v(\mathbf{x})),$$

where $v(\mathbf{x})$ is the density of $\mathbf{x}$ in $X$ defined by

$$v(\mathbf{x}) := \inf\{v > 0: \text{ for each } x \in X, \text{ there is } \text{ some } 1 \leqslant j \leqslant l \text{ such that } d(x, x_j) \leqslant \eta\}.$$

Combining the above two relations, we know that if $X \subset [-b, b]^n$, then

$$\begin{aligned} \sup_{\mathbf{x} \in X^l} \mathcal{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta) &\leqslant \mathcal{N}(\overline{I_K(B_R)}, \eta) \\ &\leqslant \sup_{\mathbf{x} \in X^l} \mathcal{N}\left(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta - R\|K\|_{C^2(X \times X)}^{1/2} \frac{b}{l^{1/n} - 1}\right). \end{aligned}$$

(The above upper bound makes sense only for $l > (R\|K\|_{C^2(X \times X)}^{1/2} b/\eta + 1)^n$.) Thus, the covering number investigated in this paper is almost equivalent to the covering numbers of discrete versions of $\overline{I_K(B_R)}$ used in some learning theory literature. Another covering number of the set $\overline{I_K(B_R)}|_{\mathbf{x}}$ is the one defined with respect to the $\ell_2$ metric: $d((a_i)_{i=1}^l, (b_i)_{i=1}^l) = (\sum_{i=1}^l |a_i - b_i|^2)^{1/2}$. This empirical $L^2$ covering number may yield better bounds and better generalization performances, which we shall not discuss here.

To estimate the covering number $\mathcal{N}(\overline{I_K(B_R)}, \eta)$, a natural idea would be to estimate the regularity of the kernel $\sum \sqrt{\lambda_j} \phi_j(x) \phi_j(t)$ of $L_K^{1/2}$. However, we shall give an example in Section 2 to show that this kernel need not be in $L^\infty(X \times X)$, hence is not necessarily a Mercer kernel.

Some previous methods for estimating the covering number in Learning Theory (for discrete versions of $I_K(B_R)$) are based on the claim that the eigenfunctions $\{\phi_j\}$ are uniformly bounded, see, *e.g.*, [12]. Steve Smale [9] observed that this claim does not hold, and he provided the method of constructing a counter example. Here we shall construct a concrete counter example of a $C^\infty$ Mercer kernel to this claim. The author is grateful to Steve Smale who suggested him to include such a counter example in this paper.

The results in [12] can be applied to a Mercer kernel *only* when the eigenfunctions of $L_K$ are uniformly bounded, *i.e.*, $C_K := \sup_j \|\phi_j\|_\infty < \infty$, see equations (39) and (40) in [12]. However, this condition is hard to check. Even for the Gaussian kernel $K(x, t) = k(x - t) = e^{-|x-t|^2/2}$ on $[-1, 1]$, it is unknown whether $C_K < \infty$ (though we expect that this is true). The counter example of $C^\infty$ kernel we shall present in Section 2 tells us that the regularity of the kernel function (or the decay of the eigenvalues of the integral operator $L_K$) does not ensure the finiteness of $C_K$.

The only class of Mercer kernels that were handled in [12] are periodic kernels. To see this, let $k(x)$ be a symmetric positive definite function on $\mathbb{R}$. Take $v > 0$ and define the periodized kernel (if $k$ has some mild decay) as

$$k_v(x) := \sum_{j=-\infty}^{+\infty} k(x - jv).$$

Expand this $v$-periodic function into its Fourier series

$$k_v(x) = \sum_{j \in \mathbb{Z}} \lambda_j \sqrt{\frac{2}{v}} \cos(2\pi jx/v).$$

Then the integral operator

$$L_K f(x) := \int_{-v/2}^{v/2} k_v(x - t) f(t) dt, \qquad x \in [-v/2, v/2]$$

can be considered to be defined on the space of square integrable $v$-periodic functions or $L^2[-v/2, v/2]$. For this operator $\{\sqrt{2/v} \cos(2\pi jx/v)\}_{j \in \mathbb{Z}}$ are eigenfunctions associated with possibly nonzero eigenvalues $\{\lambda_j\}_{j \in \mathbb{Z}}$. For this periodic kernel, we do have $C_K = \sup_j \|\phi_j\|_\infty = \sqrt{2/v} < \infty$. However, the reproducing kernel Hilbert space $\mathscr{H}_{k_v}$ induced by the kernel $K_v(x, t) := k_v(x - t)$ on $[-v/2, v/2]$ is totally different from $\mathscr{H}_k$ induced by $k(x - t)$. In fact, all the functions in $\mathscr{H}_{k_v}$ satisfy $f(-v/2) = f(v/2)$. But most fundamental functions $K_x(t) := k(x - t)$ in $\mathscr{H}_k$ do not satisfy this periodic condition. As an example, consider the Gaussian kernel $k(x) = e^{-|x|^2/2}$. The fundamental function $K_x$ satisfies $|K_x(v/2) - K_x(-v/2)| \geqslant |x| v e^{-v^2/2}$, hence the periodic condition is valid only if $x = 0$. Moreover, when $|x| \geqslant v/4$, for any $f \in \mathscr{H}_{k_v}$ we have $\|K_x - f\|_{C[-v/2, v/2]} \geqslant (v^2/4) e^{-v^2/2}$; an $\eta$-cover of $\overline{I_{K_v}(B_R)}$ cannot cover $\overline{I_K(B_R)}$ if $0 < \eta < (v^2/4) e^{-v^2/2}$ (a fixed constant). Also, the space $\mathscr{H}_{k_v}$ even does not contain any fundamental functions $K_x$ (except $K_0$) which are used as elements in empirical learning procedure. Hence the covering number for the space $\mathscr{H}_{k_v}$ cannot be applied to estimate $\mathscr{N}(\overline{I_K(B_R)}, \eta)$ or $\mathscr{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta)$.

Another simple example is the hat function $k(x)$ which is supported on $[-1, 1]$ and given by $k(x) = 1 - |x|$ for $x \in [-1, 1]$. If the period $v$ is $1/m$ for some $m \in \mathbb{N}$, then $k_v(x) \equiv m$. Hence $\mathscr{H}_{k_v}$ contains only the constant functions, while the space $\mathscr{H}_k$ has a very rich structure (it contains functions which are in Lip $\frac{1}{2}$, but not in Lip $(\frac{1}{2} + \delta)$ for any $\delta > 0$. Here for $\alpha > 0$, Lip $\alpha$ stands for the space of continuous functions $f$ satisfying $|f(x) - f(y)| \leqslant C_f |x - y|^\alpha$ for a constant $C_f$ depending on $f$ and any $x, y \in [-1, 1]$).

Similar to [12], periodic kernels were considered in [5]. The only example of nonperiodic kernel was the univariate Gaussian kernel $k(x) = e^{-x^2/\sigma^2}$. The estimate given in [5, Theorem 7] for the covering number concerning this kernel was $\ln \sup_{\mathbf{x}\in X^l} \mathcal{N}(\overline{I_K(B_R)}|_{\mathbf{x}}, \eta) \leqslant 48(\frac{R}{\eta})^2 \ln^2(\frac{4eRl}{\eta})$, which is a rough estimate.

Though the eigenfunctions of $L_K$ need not be uniformly bounded, Cucker and Smale [2] were able to show that for a $C^\infty$ Mercer kernel on $X \subset \mathbb{R}^n$, the RKHS $\mathcal{H}_K$ can be embedded into the Sobolev space $H^h$ for any $h \in \mathbb{N}$, using the results on approximation errors given by Smale and Zhou in [10]. Here the space $H^h$ consists of all the $L^2$ functions such that $D^\alpha f \in L^2$ for any $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$ with $\alpha_1 + \cdots + \alpha_n \leqslant h$. With this approach, by the rich knowledge on covering numbers of Sobolev spaces (see, *e.g.*, [3]), they proved in [2] that

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{RC_h}{\eta}\right)^{\frac{2n}{h}}, \tag{1.4}$$

where $C_h$ is a constant independent of $R$ and $\eta$.

In this paper we use a different method to estimate the covering number $\mathcal{N}(\overline{I_K(B_R)}, \eta)$. Our estimates are based on the regularity of the kernel function $K$. For convolution type kernels $K(x,t) = k(x-t)$ on $[0,1]^n$, we provide estimates depending on the decay of $\hat{k}$, the Fourier transform of $k$. In particular, when $\hat{k}$ decays exponentially, the covering number of the ball with radius $R$, $\mathcal{N}(\overline{I_K(B_R)}, \eta)$, satisfies

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant C_{k,n}\left(\ln \frac{R}{\eta}\right)^{n+1},$$

where the constant $C_{k,n}$ depends on the kernel and the dimension. This covers many important Mercer kernels. Let us show this by the example of Gaussian kernels. The estimate we give here is better than all the previous results.

Proposition 1. *Let $\sigma > 0$, $n \in \mathbb{N}$ and*

$$k(x) = \exp\left\{-\frac{|x|^2}{\sigma^2}\right\}, \qquad x \in \mathbb{R}^n.$$

*Set $X = [0,1]^n$ and the kernel $K$ as*

$$K(x,t) = k(x-t), \qquad x,t \in [0,1]^n.$$

*Then for $0 < \eta \leqslant R/2$, there holds*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(3\ln\frac{R}{\eta} + \frac{54n}{\sigma^2} + 6\right)^n \left((6n+1)\ln\frac{R}{\eta} + \frac{90n^2}{\sigma^2} + 11n + 3\right).$$

*In particular, when* $0 < \eta < R\exp\{-90n^2/\sigma^2 - 11n - 3\}$, *we have*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant 4^n(6n+2)\left(\ln \frac{R}{\eta}\right)^{n+1}.$$

## 2. MERCER KERNELS AND RELATED EIGENFUNCTIONS

In this section we present two examples of Mercer kernels on $X = [0, 1]$. The first example is a $C^\infty$ kernel, and the eigenfunctions of the integral operator (1.1) associated with this kernel are not uniformly bounded. The second example shows that the kernel $\sum \sqrt{\lambda_j}\phi_j(x)\phi_j(t)$ for $L_K^{1/2}$ need not be in $L^\infty(X \times X)$, hence is not necessarily a Mercer kernel. The construction of both examples are based on an orthonormal basis (of $L^2[0, 1]$) consisting of $C^\infty$ functions supported in $(0, 1)$ (each function is supported on $[1/N_0, 1 - 1/N_0]$ for some $N_0 \in \mathbb{N}$ depending on this function).

Lemma 2.1. *Let* $\Psi \in C^\infty(\mathbb{R})$ *be real-valued and supported on* $[1/4, 7/4]$ *such that*

$$\Psi(x) + \Psi(x+1) = 1, \qquad \forall x \in [0, 1]. \tag{2.1}$$

*Set*

$$\psi_j(x) := \frac{8^{\frac{j}{2}}}{\|\Psi\|_2} \Psi(8^j x), \qquad j \in \mathbb{N}.$$

*Then there is a sequence of* $C^\infty$ *functions* $\{\varphi_j\}_{j=1}^\infty$ *supported in* $(0, 1)$ *such that* $\{\psi_j\}_{j=1}^\infty \cup \{\varphi_j\}_{j=1}^\infty$ *form an orthonormal basis of* $L^2[0, 1]$.

*Proof.* By the support of $\Psi$ and (2.1), we know that

$$\sum_{j \in \mathbb{Z}} \Psi(x + j) \equiv 1. \tag{2.2}$$

We claim that the span of the set of functions $\{\Psi(Nx - j)\colon 1 \leqslant j \leqslant N - 2, 3 \leqslant N \in \mathbb{N}\}$ is dense in $L^2[0, 1]$.

To verify our claim, assume that $f$ is a $C^1$ function supported on $[1/N_0, 1 - 1/N_0]$ for some $N_0 \in \mathbb{N}$. Then for $N \geqslant N_0$,

$$\sum_{j \in \mathbb{Z}} f\left(\frac{j}{N}\right)\Psi(Nx - j) = \sum_{j=1}^{N-2} f\left(\frac{j}{N}\right)\Psi(Nx - j).$$

This in connection with (2.2) implies that for $x \in [0,1]$,

$$\left| f(x) - \sum_{j=1}^{N-2} f\left(\frac{j}{N}\right) \Psi(Nx-j) \right| = \left| \sum_{j \in \mathbb{Z}} \left[ f(x) - f\left(\frac{j}{N}\right) \right] \Psi(Nx-j) \right|$$

$$= \left| \sum_{Nx-7/4 \leqslant j \leqslant Nx-1/4} \left[ f(x) - f\left(\frac{j}{N}\right) \right] \Psi(Nx-j) \right| \leqslant \|f'\|_\infty \frac{7}{2N} \|\Psi\|_\infty.$$

This tells us that

$$\left\| f(x) - \sum_{j=1}^{N-2} f\left(\frac{j}{N}\right) \Psi(Nx-j) \right\|_{L^2[0,1]} \leqslant \|f'\|_\infty \frac{7}{2N} \|\Psi\|_\infty \to 0 \qquad (N \to \infty).$$

Since $C^1$ functions $f$ supported in $(0,1)$ (hence supported on $[1/N_0, 1-1/N_0]$ for some $N_0 \in \mathbb{N}$ depending on $f$) are dense in $L^2[0,1]$, the statement we claimed holds true.

Now we can construct the sequence $\{\varphi_j\}$. Since $\psi_j$ is supported on $\left[\frac{1}{4}8^{-j}, \frac{7}{4}8^{-j}\right]$, we know that $\{\psi_j\}_{j \in \mathbb{N}}$ is an orthonormal system in $L^2[0,1]$.

Note that the function $\Psi(Nx-j)$ with $1 \leqslant j \leqslant N-2, 3 \leqslant N \in \mathbb{N}$ is supported on $\left[\left(j+\frac{1}{4}\right)/N, \left(j+\frac{7}{4}\right)/N\right] \subset (0,1)$. Then by the support property, its orthogonal projection onto the orthogonal complement of $\overline{\text{span}}$ $\{\psi_l\}_{l \in \mathbb{N}}$,

$$\Psi(Nx-j) - \sum_{l=1}^{\infty} \langle \Psi(N \cdot -j), \psi_l \rangle \psi_l(x)$$

equals

$$\Psi(Nx-j) - \sum_{1 \leqslant l \leqslant \ln\left(\frac{7N}{4}\right) / \ln 8} \langle \Psi(N \cdot -j), \psi_l \rangle \psi_l(x).$$

This is a $C^\infty$ function supported in $(0,1)$. Such orthogonal projections form a countable set of $C^\infty$ functions supported in $(0,1)$ (some may be the zero function). Then we apply the Gram–Schmidt orthogonalization procedure and find an orthonormal basis

$$\{\psi_j\}_{j=1}^{\infty} \cup \{\varphi_j\}_{j=1}^{\infty}$$

of $L^2[0,1]$. The Gram-Schmidt procedure in connection with the support property tells us that each $\psi_j$ or $\varphi_j$ is a $C^\infty$ function supported in $(0,1)$. Hence Lemma 2.1 holds. ∎

We are now in a position to give the examples. Denote the norm of the Sobolev space $W_\infty^m$ as

$$\|f\|_{W_\infty^m} := \sum_{|j| \leqslant m} \|D^j f\|_{L^\infty}.$$

EXAMPLE 1. *Let* $X = [0, 1]$ *and*

$$K(x, t) := \sum_{j=1}^{\infty} \frac{2^{-j}}{\|\psi_j\|_{W_\infty^j}^2} \psi_j(x)\psi_j(t) + \sum_{j=1}^{\infty} \frac{2^{-j}}{\|\varphi_j\|_{W_\infty^j}^2} \varphi_j(x)\varphi_j(t).$$

*Then* $K$ *is a* $C^\infty$ *Mercer kernel on* $X \times X$. *The operator* $L_K$ *defined by* (1.1) *is positive and its normalized eigenfunctions are* $\psi_j$ *and* $\varphi_j$ *associated with eigenvalues* $2^{-j}/\|\psi_j\|_{W_\infty^j}^2$ *and* $2^{-j}/\|\varphi_j\|_{W_\infty^j}^2$, *respectively. However,*

$$\|\psi_j\|_{L^\infty(X)} = \left\| \frac{8^{j/2}}{\|\Psi\|_2} \Psi(8^j x) \right\|_{L^\infty} = 8^{j/2} \frac{\|\Psi\|_{L^\infty}}{\|\Psi\|_2} \to \infty \qquad (j \to \infty).$$

*Proof.* We only need to prove that $K \in C^\infty(X \times X)$. Let $\alpha, \beta \in \mathbb{Z}_+$. Then for any $s \geqslant \max\{\alpha, \beta\}$,

$$\sum_{j=s}^{\infty} \frac{2^{-j}}{\|\psi_j\|_{W_\infty^j}^2} |D^\alpha \psi_j(x)|\,|D^\beta \psi_j(t)| + \sum_{j=s}^{\infty} \frac{2^{-j}}{\|\varphi_j\|_{W_\infty^j}^2} |D^\alpha \varphi_j(x)|\,|D^\beta \varphi_j(t)|$$
$$\leqslant 2 \sum_{j=s}^{\infty} 2^{-j} = 2^{2-s}.$$

Hence the series defining $K$ converges in $C^m(X \times X)$ for any $m \in \mathbb{N}$. Therefore, $K \in C^\infty(X \times X)$. ∎

It was claimed in [12, Theorem 4] that the $L^\infty$-norms of the normalized eigenfunctions of $L_K$ associated with a Mercer kernel are bounded. Obviously, Example 1 provides a counter example to the above claim. Such a counter example in a slightly different form (without the $C^\infty$ regularity) was originally shown to the author by Steve Smale [9]. As the kernel in Example 1 is $C^\infty$, we know that the regularity of the kernel (or the decay of the eigenvalues of the integral operator) does not help to ensure that the eigenfunctions are uniformly bounded.

EXAMPLE 2. *Let* $X = [0, 1]$ *and*

$$K(x, t) := \sum_{j=1}^{\infty} \frac{\|\Psi\|_2^4}{j^2 8^j} \psi_j(x)\psi_j(t) + \sum_{j=1}^{\infty} \frac{4^{-j}}{(\|\varphi_j\|_{L^\infty}^2 + 1)^2} \varphi_j(x)\varphi_j(t).$$

*Then $K$ is a Mercer kernel on $X \times X$. However, the function*

$$K^{[1/2]}(x,t) := \sum_{j=1}^{\infty} \sqrt{\frac{\|\Psi\|_2^4}{j^2 8^j}}\,\psi_j(x)\psi_j(t) + \sum_{j=1}^{\infty} \sqrt{\frac{4^{-j}}{(\|\varphi_j\|_{L^\infty}^2+1)^2}}\,\varphi_j(x)\varphi_j(t)$$

*does not lie in $L^\infty(X \times X)$.*

*Proof.* As the series defining $K$ converges uniformly, we know that $K \in C(X \times X)$ and it is a Mercer kernel.

To see the second statement, we restrict $x, t$ to $[\frac{1}{4}8^{-m},\ \frac{7}{4}8^{-m}]$. Then

$$|K^{[1/2]}(x,t)| \geqslant \left|\frac{8^{m/2}}{m}\,\Psi(8^m x)\Psi(8^m t)\right| - \sum_{j=1}^{\infty} 2^{-j}.$$

Hence

$$\|K^{[1/2]}(x,t)\|_{L^\infty(X\times X)} \geqslant \frac{8^{m/2}}{m}\,\|\Psi\|_{L^\infty}^2 - 1 \to +\infty \qquad (m \to \infty).$$

Therefore, $K^{[1/2]} \notin L^\infty(X \times X)$. ∎

Example 2 demonstrates that for a continuous Mercer kernel, the square root $L_K^{1/2}$ of the integral operator $L_K$ (which may be used to define the RKHS $\mathscr{H}_K$) need not be defined by a Mercer kernel as (1.1). We conjecture that there exists a $C^\infty$ Mercer kernel on $[0,1]$ such that $L_K^{1/2}$ is not an integral operator defined by a Mercer kernel.

## 3. ESTIMATING THE COVERING NUMBER

In this section we estimate the covering number $\mathcal{N}(\overline{I_K(B_R)}, \eta)$ for the balls of a RKHS $\mathscr{H}_K$ as subsets of $C(X)$. To this end, we need the following function to measure the regularity of the kernel. Assume that $\{X_N\colon N \in \mathbb{N}\}$ is a family of finite subsets of $X$ such that

$$d_N := \max_{x\in X} \min_{y\in X_N} d(x,y) \to 0 \qquad (N \to \infty).$$

This means that the discrete knots $X_N$ become dense as $N$ tends to the infinity. Then the function measuring the regularity of $K$ is defined as

$$\varepsilon_K(N) := \sup_{x\in X}\left\{\inf\left\{K(x,x) - 2\sum_{y\in X_N} w_y K(x,y) + \sum_{y,t\in X_N} w_y K(y,t) w_t\colon\ w_y \in \mathbb{R}\right\}^{1/2}\right\}. \tag{3.1}$$

By choosing $w_y = \delta_{y,t}$ for some $t$ depending on $x$, we can see that $\lim_{N\to\infty} \varepsilon_K(N) = 0$. For analytic kernels, $\varepsilon_K(N)$ often decays exponentially.

As an example to see the measurement, suppose that for some $0 < s \leqslant 1$, the kernel $K$ is Lip $s$ in the sense that

$$|K(x, y) - K(x, t)| \leqslant C(d(y, t))^s, \qquad \forall x, y, t \in X,$$

where $C$ is a constant independent of $x, y, t$. Let $x \in X$. Choose $y_0 \in X_N$ such that $d(x, y_0) \leqslant d_N$. Set the coefficients $\{w_y\}_{y\in X_N}$ as

$$w_y = \begin{cases} 1, & \text{if } y = y_0, \\ 0, & \text{otherwise.} \end{cases}$$

Then the Lip $s$ regularity and the symmetry yield

$$\begin{aligned} &K(x,x) - 2\sum_{y\in X_N} w_y K(x, y) + \sum_{y,t\in X_N} w_y K(y, t) w_t \\ &= K(x,x) - 2K(x, y_0) + K(y_0, y_0) \leqslant 2C(d(x, y_0))^s \leqslant 2Cd_N^s. \end{aligned}$$

Hence

$$\varepsilon_K(N) \leqslant 2Cd_N^s, \qquad \forall N \in \mathbb{N}.$$

In particular, if $X = [0, 1]$ and $X_N = \{j/N\}_{j=0}^N$, then $d_N \leqslant 1/(2N)$, thereby $\varepsilon_K(N) \leqslant 2^{1-s} C N^{-s}$.

More details and examples for estimating (3.1) will be given after we present our main result of this section.

Denote $\# X_N$ as the cardinality of the set $X_N$, and $A_N$ be the positive definite matrix

$$A_N := [K(y, t)]_{y,t\in X_N}.$$

Then our main result can be stated as follows.

THEOREM 1. *Let $I_K$ be given as above. Then for $0 < \eta \leqslant R/2$, there holds*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant (\# X_N) \ln\left( 8\|K\|_\infty^{3/2} (\# X_N) \|A_N^{-1}\|_{l^2(X_N)} \frac{R}{\eta} \right), \tag{3.2}$$

*where $N$ is any integer satisfying*

$$\varepsilon_K(N) \leqslant \frac{\eta}{2R}. \tag{3.3}$$

*Proof.* Let $N$ satisfy (3.3). Define a set of nodal functions $\{u_y(x)\}_{y\in X_N}$ by

$$[u_y(x)]_{y\in X_N} = A_N^{-1} [K(x, t)]_{t\in X_N},$$

*i.e.*,

$$\sum_{t\in X_N} K(y,t)u_t(x) = K(x,y), \qquad y\in X_N,\ \ x\in X.$$

Let $f\in B_R$. Then by the reproducing property (1.3), for $x\in X$,

$$f(x) - \sum_{y\in X_N} f(y)u_y(x) = \left\langle K_x - \sum_{y\in X_N} u_y(x)K_y, f\right\rangle_{\mathscr{H}_K}.$$

Hence the Schwartz inequality gives

$$\left|f(x) - \sum_{y\in X_N} f(y)u_y(x)\right| \leqslant \|f\|_K \left(\left\langle K_x - \sum_{y\in X_N} u_y(x)K_y, K_x - \sum_{y\in X_N} u_y(x)K_y\right\rangle_{\mathscr{H}_K}\right)^{1/2}.$$

By (1.3), this inner product equals

$$K(x,x) - 2\sum_{y\in X_N} u_y(x)K(x,y) + \sum_{y,t\in X_N} u_y(x)K(y,t)u_t(x).$$

However, the quadratic function

$$Q((w_y)_{y\in X_N}) := K(x,x) - 2\sum_{y\in X_N} w_yK(x,y) + \sum_{y,t\in X_N} w_yK(y,t)w_t$$

over $\mathbb{R}^{X_N}$ takes its minimum value at $(u_y(x))_{y\in X_N}$. Therefore, $|f(x) - \sum_{y\in X_N} f(y)u_y(x)|$ is bounded by

$$\|f\|_K \inf\left\{K(x,x) - 2\sum_{y\in X_N} w_yK(x,y) + \sum_{y,t\in X_N} w_yK(y,t)w_t\colon\ w_y\in\mathbb{R}\right\}^{1/2}.$$

It follows from the definition (3.1) that

$$\left\|f(x) - \sum_{y\in X_N} f(y)u_y(x)\right\|_{C(X)} \leqslant R\varepsilon_K(N). \tag{3.4}$$

Observe that

$$\begin{aligned}\|[u_y(x)]_{y\in X_N}\|_{l^2(X_N)} &\leqslant \|A_N^{-1}\|_{l^2(X_N)}\|[K(x,t)]_{t\in X_N}\|_{l^2(X_N)}\\ &\leqslant \|A_N^{-1}\|_{l^2(X_N)}\sqrt{\#X_N}\|K\|_\infty.\end{aligned}$$

Then for any $(c_y), (d_y) \subset l^2(X_N)$, we have

$$\left\| \sum_{y \in X_N} c_y u_y(x) - \sum_{y \in X_N} d_y u_y(x) \right\|_{C(X)} \leqslant \|\{c_y - d_y\}\|_{l^2(X_N)} \|A_N^{-1}\|_{l^2(X_N)} \sqrt{\#X_N} \|K\|_\infty.$$

Notice from (1.3) that

$$|f(x)| \leqslant |\langle K_x, f \rangle_{\mathscr{H}_K}| \leqslant R\sqrt{K(x,x)} \leqslant R\sqrt{\|K\|_\infty}.$$

Now we can estimate the covering number.

By (3.4) and (3.3), for any $f \in B_R$, we can find $d = (f(y))_{y \in X_N} \in l^2(X_N)$ such that

$$\left\| f(x) - \sum_{y \in X_N} d_y u_y(x) \right\|_{C(X)} \leqslant \frac{\eta}{2}$$

and

$$\|d\|_{l^2(X_N)} \leqslant \sqrt{\#X_N \|K\|_\infty} R.$$

It is well known (see, *e.g.*, [2]) that for a finite dimensional space $E$ with $\dim E = m$,

$$\ln \mathscr{N}(B_r, \varepsilon) \leqslant m \ln(4r/\varepsilon).$$

Consider the space $l^2(X_N)$ with dimension $\#X_N$. Take

$$r = \sqrt{\#X_N \|K\|_\infty} R, \qquad \varepsilon = \frac{\eta/2}{\|A_N^{-1}\|_{l^2(X_N)} \sqrt{\#X_N} \|K\|_\infty}.$$

Then there are sequences $\{c^l \colon l = 1, \ldots, [(4r/\varepsilon)^{\#X_N}]\} \subset l^2(X_N)$ such that for any $d \in l^2(X_N)$ with $\|d\|_{l^2(X_N)} \leqslant r$, we can find some $l$ satisfying

$$\|d - c^l\|_{l^2(X_N)} \leqslant \varepsilon.$$

It follows that

$$\left\| \sum_{y \in X_N} c_y^l u_y(x) - \sum_{y \in X_N} d_y u_y(x) \right\|_{C(X)} \leqslant \varepsilon \|A_N^{-1}\|_{l^2(X_N)} \sqrt{\#X_N} \|K\|_\infty = \eta/2.$$

Thus,

$$\left\| f(x) - \sum_{y \in X_N} c_y^l u_y(x) \right\|_{C(X)} \leqslant \eta.$$

In this way, $I_K(B_R)$ is covered by balls with centers $\sum_{y\in X_N} c_y^l u_y(x)$ and radius $\eta$. Therefore,

$$\mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{4r}{\varepsilon}\right)^{\#X_N}.$$

That is,

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant (\#X_N) \ln\left(8\|K\|_\infty^{3/2}(\#X_N)\|A_N^{-1}\|_{l^2(X_N)}\frac{R}{\eta}\right).$$

This is the desired inequality. ∎

To show how the function (3.1) measures the regularity of the kernel $K$, we consider the univariate case: $X = [0,1]$ and $K \in C^s([0,1]^2)$ for some $s \in \mathbb{N}$. The multivariate case can be handled similarly.

Take $X_N := \{j/N\}_{j=0}^{N-1}$. Then $d_N = 1/N \to 0$.

Define univariate functions $\{w_{l,s}(t)\}_{l=0}^{s}$ as

$$w_{l,s}(t) := \sum_{j=l}^{s} \frac{st(st-1)\cdots(st-j+1)}{j!}\binom{j}{l}(-1)^{j-l}. \tag{3.5}$$

Then there holds

$$\sum_{l=0}^{s} w_{l,s}(t)z^l = \sum_{j=0}^{s} \frac{st(st-1)\cdots(st-j+1)}{j!}(z-1)^j, \qquad \forall z \in \mathbb{C}.$$

In particular,

$$\sum_{l=0}^{s} w_{l,s}(t) \equiv 1. \tag{3.6}$$

It can be easily checked that

$$w_{l,s}\left(\frac{m}{s}\right) = \delta_{l,m}, \qquad l, m \in \{0, 1, \ldots, s\}.$$

This means that $w_{l,s}$ is the Lagrange interpolation polynomial:

$$w_{l,s}(t) = \prod_{j\in\{0,1,\ldots,s\}\setminus\{l\}} \frac{t - j/s}{l/s - j/s} = \prod_{j\in\{0,1,\ldots,s\}\setminus\{l\}} \frac{st-j}{l-j}.$$

Then the norm of this function can be estimated as follows.

LEMMA 3.1. *Let $s \in \mathbb{N}, l \in \{0, 1, \ldots, s\}$. Then we have*

$$|w_{l,s}(t)| \leqslant s \binom{s}{l}, \qquad \forall t \in [0, 1].$$

*Proof.* Let $m \in \{0, 1, \ldots, s-1\}$ and $st \in (m, m+1)$. Then for $l \in \{0, 1, \ldots, m-1\}$,

$$\begin{aligned} |w_{l,s}(t)| &= \left| \prod_{j=0}^{l-1} (st - j) \prod_{j=l+1}^{m} (st - j) \prod_{j=m+1}^{s} (st - j) \right| \Big/ (l!(s-l)!) \\ &\leqslant \frac{(m+1)!(s-m)!}{(st-l)l!(s-l)!} \leqslant s \binom{s}{l}. \end{aligned}$$

When $l \in \{m+1, \ldots, s\}$,

$$\begin{aligned} |w_{l,s}(t)| &= \left| \prod_{j=0}^{m} (st - j) \prod_{j=m+1}^{l-1} (st - j) \prod_{j=l+1}^{s} (st - j) \right| \Big/ (l!(s-l)!) \\ &\leqslant \frac{(m+1)!(s-m)!}{(l-m)l!(s-l)!} \leqslant s \binom{s}{l}. \end{aligned}$$

The case $l = m$ can be seen in the same way. This proves Lemma 3.1. ∎

Using Lemma 3.1, we can estimate $\varepsilon_K(N)$ for $C^s$ kernels as follows.

PROPOSITION 2. *Let $X = [0, 1], s \in \mathbb{N}$ and $K$ be a $C^s$ Mercer kernel on $[0, 1]^2$. Then for $X_N = \{j/N\}_{j=0}^{N-1}$ and $N \geqslant s$, we have*

$$\varepsilon_K(N) \leqslant \left\{ \frac{(4s)^s(1 + s2^s)}{(s-1)!} \left\| \frac{\partial^s}{\partial y^s} K \right\|_\infty \right\} N^{-s}. \tag{3.7}$$

*Proof.* Let $x \in [0, 1]$. Then $x \in [m/N, (m+s)/N]$ for some $m \in \{0, \ldots, N-s\}$. Choose

$$w_{j/N} = \begin{cases} w_{i,s}(\frac{Nx-m}{s}), & \text{if } j = m + i \text{ and } i \in \{0, \ldots, s\}, \\ 0, & \text{otherwise.} \end{cases}$$

We can see that

$$\begin{aligned}
&K(x,x) - 2\sum_{y\in X_N} w_y K(x,y) + \sum_{y,t\in X_N} w_y K(y,t) w_t \\
&= K(x,x) - 2\sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right) \\
&K\left(x,\frac{m+i}{N}\right) + \sum_{i,j=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right) K\left(\frac{m+i}{N},\frac{m+j}{N}\right) w_{j,s}\left(\frac{Nx-m}{s}\right).
\end{aligned}$$

By (3.6) and the symmetry of $K$, this equals

$$\begin{aligned}
&\sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right)\left[K(x,x) - K\left(x,\frac{m+i}{N}\right)\right] \\
&+ \sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right)\sum_{j=0}^{s}\left[K\left(\frac{m+i}{N},x\right) - K\left(\frac{m+i}{N},\frac{m+j}{N}\right)\right] \\
&\times w_{j,s}\left(\frac{Nx-m}{s}\right).
\end{aligned}$$

By the Taylor expansion for the second variable, this is

$$\begin{aligned}
&-\sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right)\left\{\sum_{l=1}^{s-1}\frac{1}{l!}\left(\frac{\partial^l}{\partial y^l}K\right)(x,x)\left(\frac{m+i}{N}-x\right)^l\right. \\
&\left.+\frac{1}{s!}\left(\frac{\partial^s}{\partial y^s}K\right)(x,\xi_i)\left(\frac{m+i}{N}-x\right)^s\right\} \\
&-\sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right)\sum_{j=0}^{s} w_{j,s}\left(\frac{Nx-m}{s}\right)\left\{\sum_{l=1}^{s-1}\frac{1}{l!}\left(\frac{\partial^l}{\partial y^l}K\right)\left(\frac{m+i}{N},x\right)\right. \\
&\left.\times\left(\frac{m+j}{N}-x\right)^l + \frac{1}{s!}\left(\frac{\partial^s}{\partial y^s}K\right)\left(\frac{m+i}{N},\theta_j\right)\left(\frac{m+j}{N},-x\right)^s\right\}.
\end{aligned}$$

Here $\xi_i, \theta_i$ are numbers between $x$ and $(m+i)/N$.

Since $\{w_{i,s}(x)\}_{i=0}^{s}$ are Lagrange polynomials, we know that

$$\sum_{i=0}^{s} w_{i,s}\left(\frac{Nx-m}{s}\right)\left(\frac{m+i}{N}-x\right)^l = \left(\frac{m+s\frac{Nx-m}{s}}{N}-x\right)^l = 0, \quad \forall 1\leqslant l\leqslant s-1.$$

Therefore, we have

$$K(x,x) - 2\sum_{y\in X_N} w_y K(x,y) + \sum_{y,t\in X_N} w_y K(y,t) w_t \leqslant \frac{1}{s!}\left\|\frac{\partial^s}{\partial y^s}K\right\|_\infty$$

$$\sum_{i=0}^{s}\left|w_{i,s}\left(\frac{Nx-m}{s}\right)\right|\left(\frac{2s}{N}\right)^s + \sum_{i,j=0}^{s}\left|w_{i,s}\left(\frac{Nx-m}{s}\right)\right|\left|w_{j,s}\left(\frac{Nx-m}{s}\right)\right|$$

$$\frac{1}{s!}\left\|\frac{\partial^s}{\partial y^s}K\right\|_\infty\left(\frac{2s}{N}\right)^s.$$

Note that $\frac{Nx-m}{s}\in[0,1]$. By Lemma 3.1, this above quantity can be bounded by

$$\left\{\frac{1}{(s-1)!}\left\|\frac{\partial^s}{\partial y^s}K\right\|_\infty (4s)^s(1+s2^s)\right\}N^{-s}.$$

Thus, we have

$$\varepsilon_K(N)\leqslant\left\{\frac{(4s)^s(1+s2^s)}{(s-1)!}\left\|\frac{\partial^s}{\partial y^s}K\right\|_\infty\right\}N^{-s}.$$

This is exactly what we want. ∎

Applying Proposition 2 we can bound the integer $N$ satisfying (3.3). Once $N$ is fixed, we need to estimate $\|A_N^{-1}\|_{\ell^2(X_N)}$. Let us give such an estimation for a convolution-type kernel by means of the decay of the Fourier transform.

LEMMA 3.2. *Let $\rho>1$ and $k\in L^2(\mathbb{R})\cap C(\mathbb{R})$ satisfy*

$$\hat{k}(\xi)\geqslant C_0(1+|\xi|)^{-\rho}, \qquad \forall\xi\in\mathbb{R}. \tag{3.8}$$

*Then for $N\in\mathbb{N}$, $X_N=\{j/N\}_{j=0}^{N-1}$, and $A_N=(K(\frac{p}{N}-\frac{q}{N}))_{p,q=0}^{N-1}$, we have*

$$\|A_N^{-1}\|_{\ell^2(X_N)}\leqslant\left(\frac{5^\rho}{C_0}\right)N^{\rho-1}. \tag{3.9}$$

*Proof.* For a real vector $v:=(v_p)_{p=0}^{N-1}$, we have

$$v^T A_N v = \sum_{p,q=0}^{N-1} v_p k\left(\frac{p-q}{N}\right)v_q.$$

By the inverse Fourier transform, this equals

$$\sum_{p,q=0}^{N-1} v_p v_q \frac{1}{2\pi}\int_{\mathbb{R}}\hat{k}(\xi)e^{i\xi\cdot\frac{p-q}{N}}\,d\xi = \frac{1}{2\pi}\int_{\mathbb{R}}\hat{k}(\xi)\left|\sum_{p=0}^{N-1} v_p e^{i\xi\cdot\frac{p}{N}}\right|^2 d\xi.$$

Using (3.8) on $[-N\pi, N\pi]$, we obtain

$$v^T A_N v \geqslant \frac{1}{2\pi} \int_{-N\pi}^{N\pi} C_0 (1 + |\xi|)^{-\rho} \left| \sum_{p=0}^{N-1} v_p e^{i\xi \cdot \frac{p}{N}} \right|^2 d\xi$$

$$\geqslant \frac{C_0}{2\pi} (1 + N\pi)^{-\rho} N \int_{-\pi}^{\pi} \left| \sum_{p=0}^{N-1} v_p e^{i\xi \cdot \frac{p}{N}} \right|^2 d\xi.$$

It follows that

$$v^T A_N v \geqslant C_0 N (1 + N\pi)^{-\rho} \sum_{p=0}^{N-1} |v_p|^2 = C_0 N (1 + N\pi)^{-\rho} \|v\|_{\ell^2}^2.$$

This yields the bound:

$$\|A_N^{-1}\|_{\ell^2(X_N)} \leqslant \frac{(1 + N\pi)^\rho}{C_0 N} \leqslant \left( \frac{5^\rho}{C_0} \right) N^{\rho - 1}.$$

Thus (3.9) holds. ∎

Now we can present an example to show how to apply Theorem 1 in connection with Proposition 2 and Lemma 3.2.

PROPOSITION 3. *Let* $s \in \mathbb{N}$, $\rho > 1$ *and* $k \in L^2(\mathbb{R})$ *satisfy* (3.8). *Suppose that* $k \in C^s[-1, 1]$. *Set* $K(x, t) = k(x - t)$ *to be the Mercer kernel on* $[0, 1]^2$. *Denote a positive constant* $C_1 \geqslant ((4s)^s (1 + s2^s)/(s - 1)!) \|k^{(s)}\|_{L^\infty[-1,1]}$. *Then for* $0 < \eta \leqslant 2 s^{-s} C_1 R$ *we have*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant 2^{1+\frac{1}{s}} C_1^{\frac{1}{s}} \left( \frac{R}{\eta} \right)^{\frac{1}{s}} \left\{ \left( 1 + \frac{\rho}{s} \right) \ln \left( \frac{R}{\eta} \right) + \ln \left( 8 \|k\|_{L^\infty[-1,1]}^{3/2} \frac{5^\rho 2^{(1+\frac{1}{s})\rho} C_1^{\frac{\rho}{s}}}{C_0} \right) \right\}.$$

*Proof.* As in Proposition 2, we take $X = [0, 1]$ and $X_N = \{\frac{j}{N}\}_{j=0}^{N-1}$. By Proposition 2, for $N \geqslant s$ there holds

$$\varepsilon_K(N) \leqslant \left\{ \frac{(4s)^s (1 + s2^s)}{(s - 1)!} \|k^{(s)}\|_{L^\infty[-1,1]} \right\} N^{-s}.$$

Choose $N \geqslant (2RC_1/\eta)^{1/s}$. Then $N \geqslant s$. Since $2RC_1/\eta \geqslant 1$, we can choose such an integer satisfying further $N \leqslant 2(2RC_1/\eta)^{1/s}$. Then (3.3) holds.

Once $N$ is fixed, the norm of the inverse matrix of $A_N$ can be bounded by (3.9). Then Theorem 1 tells us that

$$\begin{aligned}\ln \mathscr{N}(\overline{I_K(B_R)}, \eta) &\leqslant N \ln\left(8\|k\|^{3/2}_{L^\infty[-1,1]} \frac{5^\rho}{C_0} N^\rho \frac{R}{\eta}\right) \\ &\leqslant 2^{1+\frac{1}{s}} C_1^{\frac{1}{s}} \left(\frac{R}{\eta}\right)^{\frac{1}{s}} \left\{ \left(1+\frac{\rho}{s}\right) \ln\left(\frac{R}{\eta}\right) \right. \\ &\qquad \left. + \ln\left(8\|k\|^{3/2}_{L^\infty[-1,1]} \frac{5^\rho 2^{(1+\frac{1}{s})\rho} C_1^{\frac{\rho}{s}}}{C_0}\right)\right\}.\end{aligned}$$

Hence the conclusion in Proposition 3 holds. ∎

Concrete examples satisfying (3.8) can be easily constructed. For example, in the construction of approximants in [10], the following kernel function appeared:

$$\varphi_0(x) = \begin{cases} 1 - \frac{3}{2}|x| + \frac{1}{2}|x|^3, & \text{if } |x| < 1, \\ 0, & \text{otherwise.} \end{cases}$$

From this kernel we can derive a family of kernels satisfying the lower bound (3.8).

EXAMPLE 3. *Let $2 \leqslant m \in \mathbb{N}$ and $k_m$ be the m-fold convolution of $\varphi_0$:*

$$k_1 = \varphi_0, \qquad k_{l+1}(x) = k_l * \varphi_0, \ \ l \in \mathbb{N}.$$

*Then $k_m(x-y)$ is a Mercer kernel on $[0,1]^2$. The low bound (3.8) holds with $C_0 = 4^{-m}$ and $\rho = 2m$. The kernel $k_m$ is $C^{2m-2}$, and there holds for $0 < \eta < 4(8\pi)^{2m-2}R$,*

$$\begin{aligned}\ln \mathscr{N}(\overline{I_K(B_R)}, \eta) \leqslant\ & 64\pi(m-1)\left(\frac{R}{\eta}\right)^{\frac{1}{2m-2}} \\ & \left\{3\ln\left(\frac{R}{\eta}\right) + 2m\ln 20 + 10\ln 2 + \ln((m-1)\pi)\right\}.\end{aligned}$$

*Proof.* The Fourier transform of $\varphi_0$ is given by [10]

$$\hat{\varphi}_0(\xi) = \frac{3}{\xi^2}\left\{1 - \frac{2\sin\xi}{\xi} + \frac{2(1-\cos\xi)}{\xi^2}\right\}.$$

Then the Fourier transform of $k_m$ satisfies

$$\hat{k}_m(\xi) \geqslant \begin{cases} 4^{-m}, & \text{if } |\xi| \leqslant \pi, \\ (3(1-2/\pi))^m |\xi|^{-2m}, & \text{if } |\xi| > \pi. \end{cases}$$

It follows that $k_m(x-y)$ is a Mercer kernel and (3.8) holds with $C_0 = 4^{-m}$ and $\rho = 2m$.

Also, since $|\hat{k}_m(\xi)| = O(|\xi|^{-2m})$, we know that $k_m \in C^{2m-2}$. Moreover,

$$\begin{aligned}\|k_m^{(2m-2)}\|_{L^\infty(\mathbb{R})} &\leqslant \frac{1}{2\pi}\int_{\mathbb{R}} |\xi|^{2m-2}|\hat{k}_m(\xi)|d\xi \\ &\leqslant \frac{1}{2\pi}\left\{\int_{-\pi}^{\pi} |\xi|^{2m-2} d\xi + \int_{|\xi|>\pi} 3^m|\xi|^{-2}(1+2/\pi+4/\pi^2)^m d\xi\right\} \\ &\leqslant \frac{\pi^{2m-2}}{2}.\end{aligned}$$

Take $C_1 = 2(16\pi(m-1))^{2m-2}$. Then the estimate for the covering number follows from Proposition 3. ∎

## 4. CONVOLUTION TYPE KERNELS

Now we turn to the special setting. Assume that $k$ is a symmetric function in $L^2(\mathbb{R}^n)$ and $\hat{k}(\xi) > 0$ almost everywhere on $\mathbb{R}^n$. Set $K$ to be the **convolution type kernel** given by $K(x,t) = k(x-t)$. Let $X = [0,1]^n$. Then $K$ is a Mercer kernel. Because of the special structure of the convolution type kernels, we may give an estimate for the covering number without using the norm of the inverse matrix $\|A_N^{-1}\|$. To this end, we need some preliminary estimates about the Taylor series expansion of the function $z^{Nx}$.

We need the Lagrange interpolation polynomial $\{w_{j,s}(t)\}_{j=0}^{s}$ defined by (3.5) again with s replaced by $N$.

Denote $X_N := \{0,1,\ldots,N\}^n$. The multivariate polynomials $\{w_{\alpha,N}(x)\}_{\alpha\in X_N}$ are defined as

$$w_{\alpha,N}(x) = \prod_{j=1}^{n} w_{\alpha_j,N}(x_j), \qquad x = (x_1,\ldots,x_n), \quad \alpha = (\alpha_1,\ldots,\alpha_n). \quad (4.1)$$

For these polynomials, we have

Lemma 4.1. *Let $x \in [0,1]^n$ and $N \in \mathbb{N}$. Then*

$$\sum_{\alpha\in X_N} |w_{\alpha,N}(x)| \leqslant (N2^N)^n \quad (4.2)$$

*and for $\theta \in [-\frac{1}{2},\frac{1}{2}]^n$, there holds*

$$\left|e^{-i\theta\cdot Nx} - \sum_{\alpha\in X_N} w_{\alpha,N}(x)e^{-i\theta\cdot\alpha}\right| \leqslant n\left(1+\frac{1}{2^N}\right)^{n-1}\left(\max_{1\leqslant j<n} |\theta_j|\right)^N. \quad (4.3)$$

*Proof.* The bound (4.2) follows directly from Lemma 3.1.

To derive the second bound (4.3), we first consider the univariate case.

Let $t \in [0,1]$. Then the univariate function $z^{Nt}$ is analytic on the region $|z-1| \leqslant 1/2$ when we take the branch such that $1^{Nt} = 1$. On this region,

$$z^{Nt} + (1 + (z-1))^{Nt} = \sum_{j=0}^{\infty} \frac{Nt(Nt-1)\cdots(Nt-j+1)}{j!}(z-1)^j.$$

It follows that for $\eta \in [-\frac{1}{2}, \frac{1}{2}]$ and $z = e^{-i\eta}$,

$$\begin{aligned}
&\left| e^{-i\eta\cdot Nt} - \sum_{j=0}^{N} \frac{Nt(Nt-1)\cdots(Nt-j+1)}{j!}(e^{-i\eta}-1)^j \right| \\
&= \left| e^{-i\eta\cdot Nt} - \sum_{l=0}^{N} w_{l,N}(t) e^{-i\eta\cdot l} \right| \\
&\leqslant \sum_{j=N+1}^{\infty} \left| \frac{Nt(Nt-1)\cdots(Nt-j+1)}{j!} \right| |\eta|^j \leqslant |\eta|^N.
\end{aligned}$$

Hence

$$\left| \sum_{l=0}^{N} w_{l,N}(t) e^{-i\eta\cdot l} \right| \leqslant 1 + |\eta|^N \leqslant 1 + \frac{1}{2^N}.$$

Now we can derive the bound in the multivariate case. Let $\theta \in [-\frac{1}{2}, \frac{1}{2}]^n$. Then

$$\begin{aligned}
&\left| e^{-i\theta\cdot Nx} - \sum_{\alpha\in X_N} w_{\alpha,N}(x) e^{-i\theta\cdot\alpha} \right| \\
&= \left| \sum_{m=1}^{n} \left[ \prod_{s=1}^{m-1} e^{-i\theta_s\cdot Nx_s} \right] \left[ e^{-i\theta_m\cdot Nx_m} - \sum_{\alpha_m=0}^{N} w_{\alpha_m,N}(x_m) e^{-i\theta_m\cdot\alpha_m} \right] \right. \\
&\qquad \left. \times \prod_{s=m+1}^{n} \left[ \sum_{\alpha_s=0}^{N} w_{\alpha_s,N}(x_s) e^{-i\theta_s\cdot\alpha_s} \right] \right|.
\end{aligned}$$

Applying the estimate in the univariate case, we see that the above expression can be bounded by

$$\sum_{m=1}^{n} \left( \max_{1\leqslant j\leqslant n} |\theta_j| \right)^N \left( 1 + \frac{1}{2^N} \right)^{n-m} \leqslant n \left( 1 + \frac{1}{2^N} \right)^{n-1} \left( \max_{1\leqslant j\leqslant n} |\theta_j| \right)^N.$$

Thus, the bound (4.3) has been verified. ∎

We are in a position to state the estimates on the covering number $\mathcal{N}(\overline{I_K(B_R)}, \eta)$ for a convolution type kernel $K(x,t) = k(x-t)$. The following function measures the regularity of the kernel function $k$:

$$\lambda_k(N) := n\left(1+\frac{1}{2^N}\right)^{n-1} \max_{1\leqslant j\leqslant n} \left\{(2\pi)^{-n}\int_{\xi\in[-\frac{N}{2},\frac{N}{2}]^n} \hat{k}(\xi)\left(\frac{|\xi_j|}{N}\right)^N d\xi\right\} + (1+(N2^N)^n)^2(2\pi)^{-n}\int_{\xi\notin[-\frac{N}{2},\frac{N}{2}]^n} \hat{k}(\xi)d\xi. \tag{4.4}$$

This function involves two parts. The first part is $\xi\in[-\frac{N}{2},\frac{N}{2}]^n$, where $(\frac{|\xi_j|}{N})^N \leqslant 2^{-N}$, hence it decays exponentially fast as $N$ becomes large. The second part is $\xi\notin[-\frac{N}{2},\frac{N}{2}]^n$, where $\xi$ is large. Then the decay of $\hat{k}$ (which is equivalent to the regularity of $k$) yields the fast decay of the second part. For more details and examples of bounding $\lambda_k(N)$ by means of the decay of $\hat{k}$ (or equivalently, the regularity of $k$), see Example 4 (Gaussian kernels) and Theorem 3 (kernels with exponentially decaying Fourier transforms) later in this section.

Theorem 2. *Assume that $k$ is a symmetric function in $L^2(\mathbb{R}^n)$* and $\hat{k}(\xi)>0$ *almost everywhere on $\mathbb{R}^n$. Let $K(x,t)=k(x-t)$ for $x,t\in[0,1]^n$. Suppose* $\lim_{N\to\infty}\lambda_k(N)=0$. *Then for $0<\eta<R/2$, there holds*

$$\ln\mathcal{N}(\overline{I_K(B_R)},\eta)\leqslant(N+1)^n\ln\left(8\sqrt{k(0)}(N+1)^{n/2}(N2^N)^n\frac{R}{\eta}\right), \tag{4.5}$$

*where $N$ is any integer satisfying*

$$\lambda_k(N)\leqslant\left(\frac{\eta}{2R}\right)^2. \tag{4.6}$$

*Proof.* Let $f\in B_R$. As in the proof of Theorem 1, for $x\in[0,1]^n$, we have

$$\left|f(x)-\sum_{\alpha\in X_N} f\left(\frac{\alpha}{N}\right)w_{\alpha,N}(x)\right|\leqslant\|f\|_K\{Q_N(x)\}^{1/2},$$

where

$$Q_N(x) := k(0)-2\sum_{\alpha\in X_N} w_{\alpha,N}(x)k\left(x-\frac{\alpha}{N}\right)+\sum_{\alpha,\beta\in X_N} w_{\alpha,N}(x)k\left(x-\frac{\alpha-\beta}{N}\right)w_{\beta,N}(x).$$

By the inverse Fourier transform and the symmetry,

$$\begin{aligned} Q_N(x) &= (2\pi)^{-n} \int_{\mathbb{R}^n} \hat{k}(\xi) \left| 1 - \sum_{\alpha \in X_N} w_{\alpha,N}(x) e^{i\xi\cdot(x-\frac{\alpha}{N})} \right|^2 d\xi \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} \hat{k}(\xi) \left| e^{-i\frac{\xi}{N}\cdot Nx} - \sum_{\alpha \in X_N} w_{\alpha,N}(x) e^{-i\frac{\xi}{N}\cdot\alpha} \right|^2 d\xi. \end{aligned}$$

Now we separate this integral into two parts, one with $\xi \in [-\frac{N}{2}, \frac{N}{2}]^n$ and the other with $\xi \notin [-\frac{N}{2}, \frac{N}{2}]^n$. For the first region, (4.3) in Lemma 4.1 with $\theta = \frac{\xi}{N}$ tells us that

$$\begin{aligned} &\int_{\xi \in [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) \left| e^{-i\frac{\xi}{N}\cdot Nx} - \sum_{\alpha \in X_N} w_{\alpha,N}(x) e^{-i\frac{\xi}{N}\cdot\alpha} \right|^2 d\xi \\ &\leqslant n\left(1 + \frac{1}{2^N}\right)^{n-1} \max_{1 \leqslant j \leqslant n} \int_{\xi \in [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) \left(\frac{|\xi_j|}{N}\right)^N d\xi. \end{aligned}$$

For the second region, we apply (4.2) in Lemma 4.1 and obtain

$$\begin{aligned} &\int_{\xi \notin [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) \left| e^{-i\frac{\xi}{N}\cdot Nx} - \sum_{\alpha \in X_N} w_{\alpha,N}(x) e^{-i\frac{\xi}{N}\cdot\alpha} \right|^2 d\xi \\ &\leqslant (1 + (N2^N)^n)^2 \int_{\xi \notin [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) d\xi. \end{aligned}$$

Combining the above two cases, we have

$$\begin{aligned} Q_N(x) \leqslant\ & n\left(1 + \frac{1}{2^N}\right)^{n-1} \max_{1 \leqslant j \leqslant n} \left\{ (2\pi)^{-n} \int_{\xi \in [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) \left(\frac{|\xi_j|}{N}\right)^N d\xi \right\} \\ &+ \frac{(1 + (N2^N)^n)^2}{(2\pi)^n} \int_{\xi \notin [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) d\xi = \lambda_k(N). \end{aligned}$$

Since $N$ satisfies (4.6), we know that

$$\left\| f(x) - \sum_{\alpha \in X_N} f\left(\frac{\alpha}{N}\right) w_{\alpha,N}(x) \right\|_{C(X)} \leqslant \frac{\eta}{2}.$$

Also, by (1.3),

$$\left\| \left\{ f\left(\frac{\alpha}{N}\right) \right\} \right\|_{l^2(X_N)} \leqslant R\sqrt{k(0)}(N+1)^{n/2}.$$

As in the proof of Theorem 1, we take

$$r = R\sqrt{k(0)}(N+1)^{n/2}, \qquad \varepsilon = \frac{\eta}{2(N2^N)^n}.$$

Then there are $\{c^l : l = 1, \ldots, [(4r/\varepsilon)^{(N+1)^n}]\} \subset l^2(X_N)$ such that for any $d \in l^2(X_N)$ with $\|d\|_{l^2(X_N)} \leqslant r$, we can find some $l$ satisfying

$$\|d - c^l\|_{l^2(X_N)} \leqslant \varepsilon.$$

This in connection with Lemma 4.1 yields

$$\left\| \sum_{\alpha \in X_N} c_\alpha^l w_{\alpha,N}(x) - \sum_{\alpha \in X_N} d_\alpha w_{\alpha,N}(x)) \right\|_C (X) \leqslant \eta/2.$$

Thus,

$$\left\| f(x) - \sum_{\alpha \in X_N} c_\alpha^l w_{\alpha,N}(x) \right\|_C (X) \leqslant \eta.$$

We have covered $I_K(B_R)$ by balls with centers $\sum_{\alpha \in X_N} c_\alpha^l w_{\alpha,N}(x)$ and radius $\eta$. Therefore,

$$\mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{4r}{\varepsilon}\right)^{\# X_N}.$$

That is,

$$\begin{aligned} \ln \mathcal{N}(\overline{I_K(B_R)}, \eta) &\leqslant (N+1)^n \ln\left(\frac{4r}{\eta}\right) \\ &\leqslant (N+1)^n \ln\left(8\sqrt{k(0)}(N+1)^{n/2}(N2^N)^n \frac{R}{\eta}\right). \end{aligned}$$

The proof of Theorem 2 is complete. ∎

The idea of using the function $Q_N$ in our proof has been used in the literature of radial basis functions, known as the power function. See, *e.g.*, [8, 13].

To see how to handle the function $\lambda_k(N)$ measuring the regularity of the kernel, and then to estimate the covering number, we turn to the example of Gaussian kernels.

Example 4. *Let* $\sigma > 0$ *and*

$$k(x) = \exp\left\{-\frac{|x|^2}{\sigma^2}\right\}, \qquad x \in \mathbb{R}^n.$$

*Then for* $0 < \eta < R/2$, *there holds*

$$\ln \mathscr{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(3 \ln \frac{R}{\eta} + \frac{54n}{\sigma^2} + 6\right)^n \times \left((6n+1) \ln \frac{R}{\eta} + \frac{90n^2}{\sigma^2} + 11n + 3\right). \tag{4.7}$$

*In particular, when* $0 < \eta < R \exp\left\{-\frac{90n^2}{\sigma^2} - 11n - 3\right\}$, *we have*

$$\ln \mathscr{N}(\overline{I_K(B_R)}, \eta) \leqslant 4^n (6n+2) \left(\ln \frac{R}{\eta}\right)^{n+1}. \tag{4.8}$$

*Proof.* It is well known that

$$\hat{k}(\xi) = (\sigma\sqrt{\pi})^n e^{-\frac{\sigma^2|\xi|^2}{4}}.$$

Hence $\hat{k}(\xi) > 0$ for any $\xi \in \mathbb{R}^n$.

Let us estimate the function in (4.4). For the first part with $1 \leqslant j \leqslant n$, we have

$$(2\pi)^{-n} \int_{\xi \in \left[-\frac{N}{2}, \frac{N}{2}\right]^n} (\sigma\sqrt{\pi})^n e^{-\frac{\sigma^2|\xi|^2}{4}} \left(\frac{|\xi_j|}{N}\right)^N d\xi$$
$$\leqslant \frac{\sigma\sqrt{\pi}}{2\pi} \int_{-N/2}^{N/2} e^{-\frac{\sigma^2|\xi_j|^2}{4}} \left(\frac{|\xi_j|}{N}\right)^N d\xi_j \leqslant \frac{2}{\sqrt{\pi}} \left(\frac{2}{\sigma N}\right)^N \Gamma\left(\frac{N+1}{2}\right).$$

Applying the Stirling's formula, we know that this term can be bounded by

$$2\left(\frac{2}{\sigma N}\right)^N \left(\frac{N+1}{2e}\right)^{\frac{N+1}{2}} \frac{1}{\sqrt{N+1}} e^{\frac{1}{6(N+1)}} \leqslant 2\left(\frac{\sqrt{2}}{\sigma\sqrt{eN}}\right)^N.$$

As for the second term, we have

$$(2\pi)^{-n} \int_{\xi \notin \left[-\frac{N}{2}, \frac{N}{2}\right]^n} (\sigma\sqrt{\pi})^n e^{-\frac{\sigma^2|\xi|^2}{4}} d\xi$$
$$\leqslant \frac{\sigma\sqrt{\pi}}{2\pi} \sum_{j=1}^n \int_{\xi_j \notin [-N/2, N/2]} e^{-\frac{\sigma^2|\xi_j|^2}{4}} d\xi_j$$
$$\leqslant \frac{n\sigma\sqrt{\pi}}{\pi} \int_{N/2}^{+\infty} e^{-\frac{\sigma^2}{4}(t^2 - t/2)} e^{-\frac{\sigma^2}{4}\cdot\frac{t}{2}} dt$$
$$\leqslant \frac{n\sigma}{\sqrt{\pi}} e^{-\frac{\sigma^2 N(N-1)}{16}} \frac{8}{\sigma^2} e^{-\frac{\sigma^2 N}{16}} = \frac{8n}{\sigma\sqrt{\pi}} e^{-\frac{\sigma^2}{16}N^2}.$$

Combining the above two estimates, the function $\lambda_k$ satisfies

$$\lambda_k(N) \leqslant 2n\left(1+\frac{1}{2^N}\right)^{n-1}\left(\frac{\sqrt{2}}{\sigma\sqrt{eN}}\right)^N + \frac{8n}{\sigma\sqrt{\pi}}(1+(N2^N)^n)^2 e^{-\frac{\sigma^2}{16}N^2}.$$

Notice that when $N \geqslant n+3$,

$$\left(1+\frac{1}{2^N}\right)^{n-1} \leqslant \left(1+\frac{1}{2n}\right)^{n-1} \leqslant \sqrt{e}$$

and

$$(1+(N2^N)^n)^2 \leqslant 2^{1-2n+4Nn}.$$

It follows that

$$\lambda_k(N) \leqslant 2n\sqrt{e}\left(\frac{2}{\sigma^2 eN}\right)^{N/2} + \frac{n4^{2-n}}{\sigma\sqrt{\pi}} e^{-\frac{\sigma^2}{16}N^2 + 4nN\ln 2}.$$

Choose $N \geqslant \frac{80n\ln 2}{\sigma^2}$. Then

$$\lambda_k(N) \leqslant 2n\sqrt{e}\left(\frac{1}{16en\ln 2}\right)^{N/2} + \frac{4}{\sigma\sqrt{\pi}} 2^{-nN} \leqslant 2\sqrt{e}\left(\frac{1}{16n}\right)^{N/2} + \frac{4}{\sigma\sqrt{\pi}} 2^{-nN}.$$

When $N \geqslant 2\ln\frac{R}{\eta} + 5/2$ and $N \geqslant \frac{3}{n}\ln\frac{R}{\eta} + 5/n - \ln(\sigma\sqrt{\pi})/(n\ln 2)$, we know that each term in the above estimates for $\lambda_k$ is bounded by $(\frac{\eta}{2R})^2/2$. Hence (4.6) holds.

Finally, we choose the minimal $N$ satisfying

$$N \geqslant \frac{80n\ln 2}{\sigma^2} + 3\ln\frac{R}{\eta} + 5.$$

Then (4.6) is valid for any $0<\eta<R/2$, by checking the cases $\sigma \geqslant 1$ and $\sigma<1$. By Theorem 2,

$$\begin{aligned} \ln \mathcal{N}(\overline{I_K(B_R)}, \eta) &\leqslant \left(3\ln\frac{R}{\eta} + \frac{80n\ln 2}{\sigma^2} + 6\right)^n \left(\left(\frac{5}{2}\ln 2\right)nN + \ln\frac{R}{\eta} + \ln 8\right) \\ &\leqslant \left(3\ln\frac{R}{\eta} + \frac{54n}{\sigma^2} + 6\right)^n \left((6n+1)\ln\frac{R}{\eta} + \frac{90n^2}{\sigma^2} + 11n + 3\right). \end{aligned}$$

This proves (4.7).

When $0<\eta<Re^{-\frac{90n^2}{\sigma^2}-11n-3}$, we have

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant 4^n(6n+2)\left(\ln\frac{R}{\eta}\right)^{n+1}.$$

This yields the desired estimate for the Gaussian kernels. ∎

The estimates for the covering number given in Example 4 verify the conclusion stated in Proposition 1. We conjecture that for the Gaussian kernels, there holds

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) = O\left(\ln \frac{R}{\eta}\right)^{n/2+1}.$$

As a corollary of Theorem 2 we consider kernels with exponentially decaying Fourier transforms.

THEOREM 3. *Let k as in Theorem 2 and*

$$\hat{k}(\xi) \leqslant C_0 e^{-\lambda|\xi|}, \qquad \forall \xi \in \mathbb{R}^n$$

*for some constants* $C_0 > 0$ *and* $\lambda > 4 + 2n \ln 4$. *Denote* $\Lambda := \max\{\frac{1}{e\lambda}, \frac{4^n}{e^{\lambda/2}}\}$. *Then for* $0 < \eta \leqslant 4R\sqrt{C_0}\Lambda^{(4n^2-1)/4}$, *there holds*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{4}{\ln \frac{1}{\Lambda}} \ln \frac{R}{\eta} + 1 + C_1\right)^n \left\{\left(\frac{4n}{\ln \frac{1}{\Lambda}} + 1\right) \ln \frac{R}{\eta} + C_2\right\}, \tag{4.9}$$

*where*

$$C_1 := 1 + \frac{2\ln(16C_0)}{\ln \frac{1}{\Lambda}}, \qquad C_2 := \ln\left(8\sqrt{\frac{C_0}{\lambda}}\, 2^{n/2} e^{nC_1}\right).$$

*Proof.* Let $N \in \mathbb{N}$ and $1 \leqslant j \leqslant n$. Then

$$\int_{\xi \in [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) \left(\frac{|\xi_j|}{N}\right)^N d\xi \leqslant \frac{C_0}{N^N} N^{n-1} \int_{\xi_j \in [-\frac{N}{2}, \frac{N}{2}]} |\xi_j|^N e^{-\lambda|\xi_j|} d\xi_j$$

$$\leqslant \frac{2C_0}{\lambda^{N+1} N^N} N^{n-1} N!.$$

By Stirling's formula, this is bounded by

$$2C_0 \sqrt{2\pi}\, 2^{\frac{1}{12}+1} \frac{N^{n-1/2}}{(e\lambda)^{N+1}}.$$

For the other term in (4.4), we have

$$(N2^N)^{2n} \int_{\xi \notin [-\frac{N}{2}, \frac{N}{2}]^n} \hat{k}(\xi) d\xi \leqslant C_0 (N2^N)^{2n} \int_{|\xi| \geqslant N/2} e^{-\lambda|\xi|} d\xi$$

$$= C_0 (N2^N)^{2n} \int_{N/2}^{\infty} \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} e^{-\lambda r} dr \leqslant \frac{2C_0 \pi^{n/2}}{\Gamma(\frac{n}{2})} N^{2n} \left(\frac{4n}{e^{\lambda/2}}\right)^N.$$

Combining the above two bounds, we have

$$\lambda_k(N) \leqslant 4C_0 N^{n-1/2} \left(\frac{1}{e\lambda}\right)^{N+1} + 4C_0 N^{2n} \left(\frac{4^n}{e^{\lambda/2}}\right)^N \leqslant 4C_0 N^{2n} \Lambda^N.$$

Note that $\Lambda < e^{-2}$. Then for $N \geqslant 4n^2$,

$$\lambda_k(N) \leqslant 4C_0 \Lambda^{N/2}.$$

Thus, for $0 < \eta \leqslant 4R\sqrt{C_0} \Lambda^{(4n^2-1)/4}$ we may take $2 \leqslant N \in \mathbb{N}$ such that $N \geqslant 4n^2$ and

$$4C_0 \Lambda^{N/2} \leqslant \left(\frac{\eta}{2R}\right)^2 \leqslant 4C_0 \Lambda^{(N-1)/2}. \tag{4.10}$$

Under this choice, (4.6) holds. Then by Theorem 2,

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant (N+1)^n \ln\left(8\sqrt{k(0)}(N+1)^{n/2}(N2^N)^n \frac{R}{\eta}\right).$$

Now (4.10) tells us that

$$N \leqslant 1 + \frac{2\ln(16C_0)}{\ln \frac{1}{\Lambda}} + \frac{4}{\ln \frac{1}{\Lambda}} \ln \frac{R}{\eta}.$$

Since

$$(N+1)^{n/2}(N2^N)^n \leqslant 2^{n/2} e^{nN},$$

we have

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{4}{\ln \frac{1}{\Lambda}} \ln \frac{R}{\eta} + 2 + \frac{2 \ln(16 C_0)}{\ln \frac{1}{\Lambda}}\right)^n$$

$$\times \ln \left(8 \sqrt{k(0)} 2^{n/2} e^{n(1+\frac{2\ln(16C_0)}{\ln\frac{1}{\Lambda}})} \left(\frac{R}{\eta}\right)^{\frac{4n}{\ln\frac{1}{\Lambda}}+1}\right).$$

Observe that

$$|k(0)| = |(2\pi)^{-n} \int \hat{k}(\xi) d\xi| \leqslant (2\pi)^{-n} \int C_0 e^{-\lambda|\xi|} d\xi \leqslant \frac{C_0}{\lambda}.$$

Then (4.9) follows. ∎

The next example deals with multiquadric kernels.

EXAMPLE 5. *Let* $\sigma > 4 + 2n \ln 4, \alpha > n$, *and*

$$k(x) = (\sigma^2 + |x|^2)^{-\alpha/2}, \qquad x \in \mathbb{R}^n.$$

*Then for* $0 < \eta \leqslant 4\sqrt{C_0}\sigma^{(4n^2-1)/4}R$, *there holds*

$$\ln \mathcal{N}(\overline{I_K(B_R)}, \eta) \leqslant \left(\frac{4}{\ln \frac{1}{\Lambda}} \ln \frac{R}{\eta} + 1 + C_1\right)^n \left\{\left(\frac{4n}{\ln \frac{1}{\Lambda}} + 1\right) \ln \frac{R}{\eta} + C_2\right\},$$

*where* $\Lambda = \max\{1/e\sigma, 4^n e^{\sqrt{\sigma}}\}$, $C_0 \geqslant 1$ *is a positive constant depending only on* $\alpha$, *and* $C_1, C_2$ *are the constants defined in Theorem* 3.

*Proof.* We know that there is a positive constant $C_0 \geqslant 1$ depending only on $\alpha$ such that

$$0 < \hat{k}(\xi) \leqslant C_0 e^{-\sigma|\xi|} \qquad \forall \xi \in \mathbb{R}^n.$$

Then we can apply Theorem 3 with $\lambda = \sigma$, and the desired estimate follows from Theorem 3. ∎

The main results in this section deal with kernels with exponentially decaying Fourier transforms. When the kernel is only $C^h$ for some $h \in \mathbb{N}$, then estimates of the form (1.4) can be obtained, by taking coefficient functions $w_{\alpha,N}(x)$ for defining $Q_n$ in the proof of Theorem 2 as functions depending on the Taylor expansion of the kernel function $k$ near the origin.

## REFERENCES

1. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
2. F. Cucker and S. Smale, On the mathematical foundations of learning, *Bull. Amer. Math. Soc.* **39** (2001), 1–49.
3. D. Edmunds and H. Triebel, "Function Spaces, Entropy Numbers, Differential Operators," Cambridge Univ. Press, Cambridge, UK, 1996.
4. T. Evgeniou, M. Pontil, and T. Poggio, Regularization networks and support vector machines, *Adv. Comput. Math.* **13** (2000), 1–50.
5. Y. Guo, P. L. Bartlett, J. Shawe-Taylor, and R. C. Williamson, Covering numbers for support vector machines, *in* "Proceedings of the Twelfth Annual Conference on Computational Learning Theory," pp. 267–277, Assoc. Comput. Mac., New York, 1999.
6. P. Niyogi, "The Informational Complexity of Learning," Kluwer Academic, Dordrecht, 1998.
7. M. Pontil, A note on different covering numbers in Learning Theory, preprint, 2001.
8. R. Schaback, Reconstruction of multivariate functions from scattered data, monograph manusript, 1997.
9. S. Smale, private communications, 2000–2001.
10. S. Smale and D. X. Zhou, Estimating the approximation error in learning theory, *Anal. Appl.*, to appear.
11. V. Vapnik, "Statistical Learning Theory," Wiley, New York, 1998.
12. R. C. Williamson, A. J. Smola, and B. Schölkopf, Generalization performance of regularization networks and support vector machines via entropy numbers of compact operators, *IEEE Trans. Inform. Theory*, to appear.
13. Z. Wu and R. Schaback, Local error estimates for radial basis function interpolation of scattered data, *IMA J. Numer. Anal.* **13** (1993), 13–27.