J Fourier Anal Appl (2008) 14: 89–101
DOI 10.1007/s00041-007-9003-z

Reproducing Kernel Hilbert Spaces Associated with Analytic Translation-Invariant Mercer Kernels

Hong-Wei Sun · Ding-Xuan Zhou

Received: 27 September 2006 / Revised: 27 June 2007 / Published online: 24 January 2008

Abstract In this article we study reproducing kernel Hilbert spaces (RKHS) associated with translation-invariant Mercer kernels. Applying a special derivative reproducing property, we show that when the kernel is real analytic, every function from the RKHS is real analytic. This is used to investigate subspaces of the RKHS generated by a set of fundamental functions. The analyticity of functions from the RKHS enables us to derive some estimates for the covering numbers which form an essential part for the analysis of some algorithms in learning theory.

Keywords Reproducing kernel Hilbert space · Derivative reproducing · Translation-invariant Mercer kernel · Real analyticity · Learning theory · Covering number

Mathematics Subject Classification (2000) 68T05 · 62J02

1 Introduction

Mercer kernels form an important tool for machine learning. The associated reproducing kernel Hilbert spaces are a central topic of learning theory. Their reproducing property enables many kernel-based learning algorithms such as support vector

The work is supported by City University of Hong Kong (Project No. 7001816), and National Science Fund for Distinguished Young Scholars of China (Project No. 10529101).

H.-W. Sun (✉)
School of Science, Jinan University, Jinan 250022, China
e-mail: ss_sunhw@ujn.edu.cn

H.-W. Sun
Department of Mathematics, Beijing Normal University, Beijing 100875, China

D.-X. Zhou
Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
e-mail: mazhou@cityu.edu.hk

BIRKHÄUSER

machines and least square methods to be implemented by convex quadratic programming optimization problems in finite dimensional spaces while nice approximation properties of the RKHS (often infinite dimensional) are preserved, leading to solid mathematical foundations of efficient learning algorithms [5, 14].

Let X be a metric space. A *Mercer kernel* on X is a continuous and symmetric function $K : X \times X \to \mathbb{R}$ such that for any finite set of points $\{x_1, \dots, x_\ell\} \subset X$, the matrix $(K(x_i, x_j))_{i,j=1}^{\ell}$ is positive semidefinite.

The *reproducing kernel Hilbert space* $\mathcal{H}_K$ associated with the kernel K is defined (see [1]) to be the completion of the linear span of the set of functions $\{K_x := K(x, \cdot) : x \in X\}$ with the inner product $\langle \cdot, \cdot \rangle_K$ given by $\langle K_x, K_y \rangle_K = K(x, y)$. That is, $\langle \sum_i \alpha_i K_{x_i}, \sum_j \beta_j K_{y_j} \rangle_K = \sum_{i,j} \alpha_i \beta_j K(x_i, y_j)$. The *reproducing property* takes the form

$$\langle K_x, f \rangle_K = f(x), \quad \forall\, x \in X,\ f \in \mathcal{H}_K. \tag{1.1}$$

In some problems of feature selection [6, 10] and manifold learning [2], data of derivatives may be available [9, 18]. Then reproducing properties of function derivatives in the RKHS are desired. This problem was initiated in [19] and further studied recently in [21]. The main result there asserts for $s \in \mathbb{N}$ that $\mathcal{H}_K$ can be embedded into $C^s(X)$ if the Mercer kernel K is in $C^{2s}(X \times X)$. Thus one can estimate the error in the $C^s(X)$ metric for learning algorithms by means of bounds with the $\mathcal{H}_K$ metric [16]. Throughout the article we take X to be a compact subset of $\mathbb{R}^n$ which is the closure of its nonempty interior X^o. For $\alpha = (\alpha^1, \alpha^2, \dots, \alpha^n) \in \mathbb{Z}_+^n$, $x = (x^1, x^2, \dots, x^n) \in \mathbb{R}^n$ and $f : X \to \mathbb{R}$, we denote $|\alpha| = \sum_{j=1}^n \alpha^j$, $\alpha! = \Pi_{j=1}^n \alpha^j!$ and

$$D^\alpha f(x) = \frac{\partial^{|\alpha|}}{(\partial x^1)^{\alpha^1} \cdots (\partial x^n)^{\alpha^n}} f(x).$$

Define $C^s(X)$ to be the space of continuous functions f on X such that $f|_{X^o} \in C^s(X^o)$ and for each $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| \le s$, $D^\alpha(f|_{X^o})$ has a continuous extension to X denoted as $D^\alpha f$.

The purpose of this article is to continue the study in [19, 21] and investigate real analyticity of functions from $\mathcal{H}_K$ associated with translation-invariant Mercer kernels of the type $K(x, y) = \varphi(|x - y|^2)$ with $\varphi : \mathbb{R} \to \mathbb{R}$. Then we shall apply our results to two mathematical problems concerning learning algorithms in $\mathcal{H}_K$: One is about subspaces of the type $\mathcal{H}_{K,\overline{x}} = \overline{\operatorname{span}}\{K_x : x \in \overline{x}\}$ with $\overline{x} \subseteq X$ investigated in [15, 16], and the other is about the covering numbers for the RKHS associated with translation-invariant Mercer kernels generated by real analytic functions.

The real analyticity of functions is defined as follows.

Definition 1 For a closed subset T of $\mathbb{R}^n$, we say that a function f on T is real analytic with convergence radius $r > 0$ if for any $t \in T$, there is a sequence of real coefficients $\{a_{\alpha,t}\}_{\alpha \in \mathbb{Z}_+^n}$ such that the series $\sum_{\alpha \in \mathbb{Z}_+^n} \frac{a_{\alpha,t}}{\alpha!}(x - t)^\alpha$ converges in $U(t, r) := \{x \in \mathbb{R}^n : |x - t| < r\}$ and equals to $f(x)$ for $x \in U(t, r) \cap T$.

The above definition is equivalent to the existence of an analytic extension of f to the domain $\cup_{t \in T}\{z \in \mathbb{C}^n : |z - t| < r\}$. Moreover, $a_{\alpha,t} = D^\alpha f(t)$ for $\alpha \in \mathbb{Z}_+^n$. ($a_{\alpha,t} = \lim_{k \to \infty} D^\alpha f(t_k)$ when t lies in the boundary of X and $t = \lim_{k \to \infty} t_k$ with $t_k \in X^o$)

Our first main result which will be proved in the next section can be stated as follows.

Theorem 1 *Let* $D = \max_{x,y\in X} |x-y|^2$ *and* φ *be a real analytic function on* $[0, D]$ *with convergence radius* $r > 0$. *If* $K(x, y) = \varphi(|x-y|^2)$ *is a Mercer kernel on* X, *then each function* f *in* $\mathcal{H}_K$ *is real analytic on* X *with convergence radius* $\widetilde{r} = \min\{\frac{\sqrt{r}}{2n}, \sqrt{r+D} - \sqrt{D}\}$.

As a corollary of Theorem 1, we shall show in Sect. 3 that in the univariate case $n = 1$, $\mathcal{H}_{K,\overline{x}} = \mathcal{H}_K$ if $\overline{x}$ is an infinite subset of X.

A detailed analysis of the Gaussian case has been done in [17]. It is an example of a large family of Mercer kernels appearing in Theorem 1 generated by radial basis functions [8, 12].

Example 1 If $\varphi \in C[0,\infty) \cap C^\infty(0,\infty)$ and $(-1)^k \varphi^{(k)}(t) \geq 0$ for any $k \in \mathbb{Z}_+$ and $t \in (0, \infty)$, then $K(x, y) = \varphi(|x-y|^2)$ is a Mercer kernel for any $n \in \mathbb{N}$ and $X \subset \mathbb{R}^n$. Examples include $\varphi(t) = \exp\{-\frac{t}{2\sigma^2}\}$ with $\sigma > 0$ and $\varphi(t) = (c^2 + t)^{-\beta}$ with $c > 0$.

Our second main result is to estimate covering numbers of balls of $\mathcal{H}_K$.

Definition 2 Let $\mathcal{F}$ be a compact metric space with metric d. For any $\eta > 0$ the covering number $\mathcal{N}(\mathcal{F}, \eta)$ is the smallest integer $\ell \in \mathbb{N}$ such that there exist $\{f_i\}_{i=1}^{\ell}$ satisfying $\min_{1\leq i\leq \ell} d(f, f_i) < \eta$ for any $f \in \mathcal{F}$.

It is well known [3] that for any $R > 0$, the ball

$$B_R := \{f \in \mathcal{H}_K : \|f\|_K \leq R\}$$

of the RKHS $\mathcal{H}_K$ with radius $R > 0$ is a compact subset of $C(X)$, which we denote as $I_K(B_R)$. The covering numbers for this compact set play an essential role in analyzing sample errors for many learning algorithms [3–5]. They have been extensively studied in [20] for some analytic kernels and in [19] for Sobolev smooth kernels. As an example, the following bounds were given.

Example 2 Let K be a Mercer kernel on $X \subset \mathbb{R}^n$ and $0 < \eta < R/2$.

(1) If K is C^s, then $\log \mathcal{N}(I_K(B_R), \eta) = O((R/\eta)^{2n/s})$.
(2) If $K(x, y) = \varphi(|x-y|^2)$ and $\varphi(t) = (c^2 + t)^{-\beta}$ with $c > 4 + 2n \log 4$, then $\log \mathcal{N}(I_K(B_R), \eta) = O((\log(R/\eta))^{n+1})$.

In the above second case, we restrict $c > 4 + 2n \log 4$ because our approach in [20] requires the Fourier transform of φ to decay exponentially fast with a large exponent. Our approach here enables us to remove this restriction and give more general bounds for the covering numbers.

In the following we often consider a real analytic function φ on $[0, D]$ with convergence radius $r > 0$, so the one-side derivative $\varphi_+^{(k)}(0)$ is well defined for each

$k \in \mathbb{Z}_+$. The real analyticity of φ at the origin means $\overline{\lim}_{k\to\infty}(|\varphi_+^{(k)}(0)|/k!)^{1/k} \leq 1/r$ which implies that for some constant $C_0 > 0$, there holds

$$\left|\frac{\varphi_+^{(k)}(0)}{k!}\right| \leq C_0\left(\frac{2}{r}\right)^k \quad \forall k \in \mathbb{Z}_+. \tag{1.2}$$

Theorem 2 *Let $D = \max_{x,y\in X}|x-y|^2$ and φ be a real analytic function on $[0, D]$ with convergence radius $r > 0$ such that (1.2) holds. If $K(x,y) = \varphi(|x-y|^2)$ is a Mercer kernel on X and $0 < \eta < R/2$, then*

$$\log\mathcal{N}(I_K(B_R),\eta) \leq \mathcal{N}\big(X,\widetilde{r}/2\big)\left(4\log\frac{R}{\eta} + 2 + 4\log\big(16\sqrt{C_0}+1\big)\right)^{n+1}, \tag{1.3}$$

where $\widetilde{r} = \min\{\frac{\sqrt{r}}{2n}, \sqrt{r+D}-\sqrt{D}\}$ and $\mathcal{N}(X,\widetilde{r}/2)$ is the covering number of X as a subset of $\mathbb{R}^n$.

Example 3 Let $c > 0$, $\beta > 0$ and $\varphi(t) = (c^2+t)^{-\beta}$. If $K(x,y) = \varphi(|x-y|^2)$, then for $0 < \eta < R/2$ there holds

$$\log\mathcal{N}(I_K(B_R),\eta) \leq \mathcal{N}\big(X,\widetilde{r}/2\big)\left(4\log\frac{R}{\eta} + 2 + 4\log\left(\frac{16(2\beta+1)^{\beta+1}}{c^\beta}+1\right)\right)^{n+1}.$$

Proof We know that φ is real analytic on $[0, D]$ with convergence radius $r = c^2 > 0$. Since $\varphi_+^{(k)}(0) = \Pi_{j=0}^{k-1}(-\beta-j)(c^2)^{-\beta-k}$, we see that

$$\left|\frac{\varphi_+^{(k)}(0)}{k!}\right| = \left(\frac{1}{c^2}\right)^{k+\beta}\Pi_{j=1}^k\frac{\beta+j-1}{j}.$$

Notice that for $j \geq \beta-1$, there holds $(\beta+j-1)/j \leq 2$. Then we have $\Pi_{j=1}^k\frac{\beta+j-1}{j} \leq 2^k\Pi_{1\leq j<\beta-1}\frac{\beta+j-1}{j} \leq 2^k(2\beta+1)^{\beta+1}$. Thus (1.2) holds with $C_0 = c^{-2\beta}(2\beta+1)^{\beta+1}$. Then our statement follows from (1.3). □

Remark 1 When X is an open subset of $\mathbb{R}^n$, $X = X^o$ and the space $C^s(X)$ has a natural definition. Our analysis holds true if we assume the uniform boundedness of the kernels and their derivatives to replace the compactness of X.

2 Analyticity of Functions in a RKHS

Before proving our main result, we need to recall some preliminary knowledge about derivative reproducing properties of $\mathcal{H}_K$. If $K \in C^{2s}(X\times X)$, for $\alpha \in \mathbb{Z}_+^n$ with $|\alpha| \leq s$, we extend α to $\mathbb{Z}_+^{2n}$ by adding zeros to the last n components and denote

$$D^\alpha K(x,y) = \frac{\partial^{|\alpha|}}{\partial(x^1)^{\alpha^1}\ldots\partial(x^n)^{\alpha^n}}K\big(x^1,\ldots,x^n,y^1,\ldots,y^n\big), \quad x,y\in X^o$$

and $D^\alpha K$ as a continuous extension of $D^\alpha(K|_{X^o\times X^o})$ to $X\times X$. For $x\in X$, denote $(D^\alpha K)_x$ as the function on X given by $(D^\alpha K)_x(y)=D^\alpha K(x,y)$.

The derivative reproducing properties of $\mathcal{H}_K$, stated in (2.1), (2.2), and (2.3) below, can be found in [13] and [11] where analytic kernels are also considered. See also the recent article [21]. So we only need to prove the last bound (2.4).

Lemma 1 (1) *Let $s\in\mathbb{N}$ and $K\in C^{2s}(X\times X)$ be a Mercer kernel. Then for any $x\in X$ and $|\alpha|\le s$, we have $(D^\alpha K)_x\in\mathcal{H}_K$ and*

$$D^\alpha f(x)=\big\langle \big(D^\alpha K\big)_x, f\big\rangle_K, \quad \forall f\in\mathcal{H}_K. \tag{2.1}$$

(2) *If K takes the form $K(x,y)=\varphi(|x-y|^2)$ with $\varphi\in C^{2s}[0,D]$ where $D=\max_{x,y\in X}|x-y|^2$, then*

$$\big(D^\alpha K\big)_x(y)=(-1)^{|\alpha|}\big(D^\alpha K_x\big)(y), \quad \forall x,y\in X, |\alpha|\le s. \tag{2.2}$$

(3) *If moreover φ is a real analytic function on $[0,D]$ with convergence radius $r>0$, then for any $\alpha\in\mathbb{Z}_+^n$, $x\in X$, we have $D^\alpha K_x\in\mathcal{H}_K$ and*

$$D^\alpha f(x)=(-1)^{|\alpha|}\big\langle f, D^\alpha K_x\big\rangle_K, \quad \forall f\in\mathcal{H}_K\,, \tag{2.3}$$

$$\big\|D^\alpha K_x\big\|_K\le 2^{|\alpha|}\sqrt{\big|\varphi_+^{(|\alpha|)}(0)\big|\alpha!}. \tag{2.4}$$

Proof To show inequality (2.4), let $\alpha\in\mathbb{Z}_+^n$ and $x\in X$. By (2.3) we have

$$\big\|D^\alpha K_x\big\|_K^2=\big\langle D^\alpha K_x, D^\alpha K_x\big\rangle_K=(-1)^{|\alpha|}D^\alpha\big(D^\alpha K_x\big)(x)=(-1)^{|\alpha|}\big(D^{2\alpha}K_x\big)(x).$$

Since φ is real analytic with convergence radius r, we know that the series $\sum_{k\in\mathbb{Z}_+}\frac{\varphi_+^{(k)}(0)}{k!}(t-0)^k$ converges and equals $\varphi(t)$ for $t\in[0,r)$. Hence,

$$K_x(y)=\varphi\big(|x-y|^2\big)=\sum_{k\in\mathbb{Z}_+}\frac{\varphi_+^{(k)}(0)}{k!}|x-y|^{2k} \quad \forall |y-x|<r.$$

Moreover, for any $\beta\in\mathbb{Z}_+^n$,

$$D^\beta(K_x)(y)=\sum_{k\in\mathbb{Z}_+}\frac{\varphi_+^{(k)}(0)}{k!}D^\beta\big(|x-\cdot|^{2k}\big)(y).$$

In particular, for $y=x$ we have

$$\begin{aligned}(-1)^{|\alpha|}\big(D^{2\alpha}K_x\big)(x)&=(-1)^{|\alpha|}\sum_{k=0}^{\infty}\frac{\varphi_+^{(k)}(0)}{k!}D^{2\alpha}\big(|x-\cdot|^{2k}\big)(x)\\&=(-1)^{|\alpha|}\frac{\varphi_+^{(|\alpha|)}(0)}{|\alpha|!}D^{2\alpha}\big(|x-\cdot|^{2|\alpha|}\big)(x)\end{aligned}$$

$$+ (-1)^{|\alpha|} \sum_{k=|\alpha|+1}^{\infty} \frac{\varphi_+^{(k)}(0)}{k!} D^{2\alpha}\big(|x-\cdot|^{2k}\big)(x).$$

Thus we obtain

$$\big\| D^\alpha K_x \big\|_K^2 = (-1)^{|\alpha|} \frac{\varphi_+^{(|\alpha|)}(0)}{\alpha!}(2\alpha)! \le \big|\varphi_+^{(|\alpha|)}(0)\big| 2^{2|\alpha|}\, \alpha!.$$

This proves (2.4) and Lemma *1*. □

Note that (2.2) follows from (2.1) by a special property of translation-invariant Mercer kernels:

$$\big(D^\alpha K\big)_x(y) = D^{(\alpha,0)}K(x,y) = (-1)^{|\alpha|} D^{(0,\alpha)} K(x,y) = (-1)^{|\alpha|}\big(D^\alpha K_x\big)(y).$$

Equation (2.2) does not hold for general Mercer kernels.

Example 4 Consider the dot product kernel $K(x,y) = x \cdot y$ on $X = [0,1]$. Take $\alpha = 1 \in \mathbb{Z}_+^1$, then $(D^\alpha K_x)(y) = x$ while $(D^\alpha K)_x(y) = y$. In particular, if $x = 0$, then $(D^\alpha K_x) \equiv 0$ while $(D^\alpha K)_x(y) = y$. Thus $(D^\alpha K_x) \neq (D^\alpha K)_x$ in general.

The computation in the proof of Lemma *1* leads to an interesting observation on the orthogonality of $\{D^\alpha K_x : \alpha \in \mathbb{Z}_+^n\}$.

Corollary 1 *Under the assumptions of Lemma* 1(3), *if* $\alpha, \beta \in \mathbb{Z}_+^n$ *are such that* $\alpha^i - \beta^i$ *is odd for some* $i \in \{1, \ldots, n\}$, *then* $\langle D^\alpha K_x, D^\beta K_x \rangle_K = 0$.

We are in a position to prove our first main result.

Proof of Theorem 1 Let $x \in X$ and $x_0 \in X$. Denote $r_1 = \sqrt{r+D} - \sqrt{D}$. Our first step is to expand the fundamental function K_x near the point x_0. Let $t \in U(x_0, r_1)$. Since $|x-x_0|^2 \in [0, D]$ and $|2\langle x-x_0, x_0-t\rangle + |x_0-t|^2| < 2r_1\sqrt{D} + r_1^2 = r$, we use the real analyticity of φ at the point $|x-x_0|^2$ and obtain

$$\begin{aligned} K_x(t) &= \varphi\big(|x-x_0|^2 + 2\langle x-x_0, x_0-t\rangle + |x_0-t|^2\big) \\ &= \sum_{k=0}^{+\infty} \frac{\varphi^{(k)}(|x-x_0|^2)}{k!}\big(2\langle x-x_0, x_0-t\rangle + |x_0-t|^2\big)^k. \end{aligned} \tag{2.5}$$

(Replace the derivative $\varphi^{(k)}(|x-x_0|^2)$ by the corresponding one-side derivative when $|x-x_0|^2 = 0$ or D.) Since the series (2.5) converges absolutely for any $t \in U(x_0, r_1)$, we know that the function K_x can be expanded as its Taylor series

$$K_x(t) = \sum_{\alpha \in \mathbb{Z}_+^n} \frac{D^\alpha K_x(x_0)}{\alpha!}(t-x_0)^\alpha \tag{2.6}$$

and the convergence radius is at least r_1.

Our second step is to establish expansions like (2.6) for a basis of $\mathcal{H}_K$. Because $\operatorname{span}\{K_x : x \in X\}$ is dense in $\mathcal{H}_K$, by the Schmidt orthonormalization we can get an orthonormal basis $\{\omega_k\}_{k\geq 1}$ of $\mathcal{H}_K$ such that each ω_k is a finite linear combination of K_x with $x \in X$. Then ω_k can be expressed as a Taylor series with convergence radius at least r_1

$$\omega_k(t) = \sum_{\alpha \in \mathbb{Z}_+^n} \frac{D^\alpha \omega_k(x_0)}{\alpha!}(t - x_0)^\alpha.$$

This in connection with (2.3) implies

$$\omega_k(t) = \sum_{\alpha \in \mathbb{Z}_+^n} (-1)^{|\alpha|} \frac{\langle \omega_k, D^\alpha K_{x_0}\rangle_K}{\alpha!}(t - x_0)^\alpha \quad \forall t \in U(x_0, r_1).$$

Our last step is to get the real analyticity of an arbitrary function in $\mathcal{H}_K$. Let $f \in \mathcal{H}_K$. Since $\{\omega_k\}_{k\geq 1}$ is an orthonormal basis of $\mathcal{H}_K$, we know that there is a sequence $\{c_k\}_{k\geq 1} \in \ell^2$ such that $f = \sum_{k\geq 1} c_k \omega_k$ in $\mathcal{H}_K$. It follows that for any $t \in U(x_0, \widetilde{r}) \cap X$ we have

$$\begin{aligned} f(t) &= \sum_{k\geq 1} c_k \omega_k(t) = \sum_{k\geq 1} c_k \sum_{\alpha \in \mathbb{Z}_+^n} (-1)^{|\alpha|} \frac{\langle \omega_k, D^\alpha K_{x_0}\rangle_K}{\alpha!}(t - x_0)^\alpha \\ &= \sum_{\alpha \in \mathbb{Z}_+^n} \sum_{k\geq 1} c_k \big\langle \omega_k, D^\alpha K_{x_0}\big\rangle_K \frac{(-1)^{|\alpha|}}{\alpha!}(t - x_0)^\alpha \\ &= \sum_{\alpha \in \mathbb{Z}_+^n} \big\langle f, D^\alpha K_{x_0}\big\rangle_K \frac{(-1)^{|\alpha|}}{\alpha!}(t - x_0)^\alpha. \end{aligned}$$

The third equality above holds because each series converges absolutely with convergence radius $\widetilde{r}$. To see this, by the Schwarz inequality we get from the identities $\|f\|_K = \|\{c_k\}\|_{\ell^2}$ and $\sum_{k\geq 1} |\langle \omega_k, D^\alpha K_{x_0}\rangle_K|^2 = \|D^\alpha K_{x_0}\|_K^2$ that

$$\begin{aligned} &\sum_{\alpha \in \mathbb{Z}_+^n} \sum_{k\geq 1} \left| c_k \big\langle \omega_k, D^\alpha K_{x_0}\big\rangle_K \frac{(-1)^{|\alpha|}}{\alpha!}(t - x_0)^\alpha \right| \\ &\quad \leq \sum_{\alpha \in \mathbb{Z}_+^n} \|\{c_k\}\|_{\ell^2} \left\{ \sum_{k\geq 1} \frac{1}{(\alpha!)^2} \left|\big\langle \omega_k, D^\alpha K_{x_0}\big\rangle_K\right|^2 |t - x_0|^{2|\alpha|} \right\}^{1/2} \\ &\quad \leq \|f\|_K \sum_{\alpha \in \mathbb{Z}_+^n} \frac{1}{\alpha!} \big\| D^\alpha K_{x_0}\big\|_K |t - x_0|^{|\alpha|}. \end{aligned}$$

By (2.4), the above expression can be bounded by

$$\|f\|_K \sum_{\alpha \in \mathbb{Z}_+^n} 2^{|\alpha|} \sqrt{\left|\varphi_+^{(|\alpha|)}(0)\right|/\alpha!}\,|t - x_0|^{|\alpha|}$$

$$= \|f\|_K \sum_{\ell=0}^{\infty} 2^\ell \sqrt{\left|\varphi_+^{(\ell)}(0)\right|/\ell!}|t-x_0|^\ell \sum_{|\alpha|=\ell} \sqrt{\ell!/\alpha!}$$

$$\le \|f\|_K \sum_{\ell=0}^{\infty} (2n)^\ell \sqrt{\left|\varphi_+^{(\ell)}(0)\right|/\ell!}|t-x_0|^\ell.$$

Each series above converges for any $t \in U(x_0, \widetilde{r})$. Therefore, every function $f \in \mathcal{H}_K$ is real analytic with convergence radius $\widetilde{r}$. The proof of Theorem 1 is complete. □

3 The Subspace $\mathcal{H}_{K,\overline{x}}$ of a RKHS

Let $\overline{x}$ be a subset of X. Consider the subspace $\mathcal{H}_{K,\overline{x}} = \overline{\text{span}}\{K_y : y \in \overline{x}\}$ of $\mathcal{H}_K$. Restricting learning processes from $\mathcal{H}_K$ onto the subspace $\mathcal{H}_{K,\overline{x}}$ may reduce the complexity of learning algorithms [15]. As a corollary of our first main result, when $X \subset \mathbb{R}$ we can prove the following special property. It is an easy consequence of the fact that an analytic function vanishing on an infinite set with an accumulation point is identically zero. For completeness we provide a proof here. Observe that a compact connected subset of $\mathbb{R}$ must be a closed interval.

Theorem 3 *Let $X = [c, d]$ for some $c < d$ and $\overline{x}$ be an infinite subset of X. Set $D = (d - c)^2$. If φ is a real analytic function on $[0, D]$ with positive convergence radius and $K(x, y) = \varphi(|x - y|^2)$ is a Mercer kernel on X, then $\mathcal{H}_{K,\overline{x}} = \mathcal{H}_K$.*

Proof Suppose to the contrary that $\mathcal{H}_{K,\overline{x}} \neq \mathcal{H}_K$. Since $\mathcal{H}_{K,\overline{x}}$ is a closed subspace, we can find a nonzero function $f \in \mathcal{H}_K$ such that $f \perp \mathcal{H}_{K,\overline{x}}$, that is, $f(x) = \langle f, K_x\rangle_K = 0$ for each $x \in \overline{x}$.

We apply Theorem 1 to the function f and can find some $r > 0$ such that f is real analytic on X with convergence radius r. Choose a sequence $c = c_1 < c_2 < \cdots < c_{\ell-1} < c_\ell = d$ such that $c_{i+1} < c_i + r$ for each $1 \le i \le \ell - 1$. For each $i \in \{1, \ldots, \ell\}$, we can find a sequences of real coefficients $\{a_{\alpha,i}\}_{\alpha\in\mathbb{Z}_+}$ such that the power series $\sum_{\alpha\in\mathbb{Z}_+} \frac{a_{\alpha,i}}{\alpha!}(x - c_i)^\alpha$ converges in $U(c_i, r)$ and equals to $f(x)$ for $x \in U(c_i, r) \cap X$.

Since $\overline{x}$ is an infinite subset of X and X is compact, we can find a point $x_0 \in X$ and a sequence $\{x_k\}_{k\in\mathbb{N}} \subseteq \overline{x}$ such that $\lim_{k\to\infty} x_k = x_0$. For some $i \in \{1, \ldots, \ell\}$, we have $x_0 \in U(c_i, r)$. Since $\lim_{k\to\infty} x_k = x_0$, there is some $N \in \mathbb{N}$ such that $x_k \in U(c_i, r)$ for $k \ge N$. This in connection with the fact $x_k \in \overline{x}$ implies that the power series $\sum_{\alpha\in\mathbb{Z}_+} \frac{a_{\alpha,i}}{\alpha!}(x - c_i)^\alpha$ vanishes at each $x_k \in U(c_i, r) \cap X$ (for $k \ge N$), hence is identically zero on its domain of convergence. This means $f(x) \equiv 0$ on $U(c_i, r) \cap X$. By considering the power series $\sum_{\alpha\in\mathbb{Z}_+} \frac{a_{\alpha,j}}{\alpha!}(x - c_j)^\alpha$ one by one for $j = i + 1, \ldots, \ell$, we know that $f(x) \equiv 0$ on $U(c_j, r) \cap X$ since it vanishes on $U(c_{j-1}, r) \cap X \cap U(c_j, r)$. This is also true for $j = i - 1, \ldots, 1$. So we can conclude that f is identically zero on X, which is a contradiction. Therefore, $\mathcal{H}_{K,\overline{x}} = \mathcal{H}_K$ and Theorem 3 is verified. □

Theorem 3 deals with the univariate case. Things are different in the multidimensional case.

Example 5 Let φ and $X \subset \mathbb{R}^n$ be as in Theorem 1. Assume $n \geq 2$. If there are $x_0, x_0' \in X^o$ and a sequence of points $\overline{x} = \{x_k\}_{k\in\mathbb{N}} \subseteq X$ satisfying $|x_0 - x_k| = |x_0' - x_k|$ for each $k \in \mathbb{N}$, then

$$\langle K_{x_0} - K_{x_0'}, K_{x_k}\rangle_K = \varphi\big(|x_0 - x_k|^2\big) - \varphi\big(|x_0' - x_k|^2\big) = 0 \quad \forall k \in \mathbb{N}.$$

Thus $K_{x_0} - K_{x_0'} \perp \mathcal{H}_{K,\overline{x}}$, and $\mathcal{H}_{K,\overline{x}} \neq \mathcal{H}_K$.

It would be interesting to study conditions on $\overline{x}$ such that $\mathcal{H}_{K,\overline{x}} = \mathcal{H}_K$. The following is a special example for the Gaussian kernels. To state the condition, we need the concept of variety.

Definition 3 ([7]) A subset A of $\mathbb{R}^n$ is called an affine variety, if it is the common zero of a collection of nonzero polynomials $f_\gamma \in \mathbb{R}[x^1, x^2, \ldots, x^n]$.

Corollary 2 *Let X be a compact connected subset of $\mathbb{R}^n$ which is the closure of its nonempty interior X^o. Let $K(x,y) = \exp\{-\frac{|x-y|^2}{\sigma^2}\}$ for some $\sigma > 0$. If $\overline{x} \subseteq X$ is a subset of an affine variety of $\mathbb{R}^n$, then $\mathcal{H}_{K,\overline{x}} \neq \mathcal{H}_K$.*

Proof Since $D^\alpha K_y(x) = p_y(x)K(x,y)$ and $p_y(x)$ is a polynomial of order $|\alpha|$, it is easy to see that for any polynomial $p \in \mathbb{R}[x^1, x^2, \ldots, x^n]$, the function pK_y lies in $\mathcal{H}_K$. If $\overline{x}$ is a subset of an affine variety of $\mathbb{R}^n$, there is a nonzero polynomial p satisfying $p(x) = 0$ for any $x \in \overline{x}$. Hence, $\langle pK_y, K_x\rangle_K = p(x)K_y(x) = 0$. That means $pK_y \perp \mathcal{H}_{K,\overline{x}}$ and $\mathcal{H}_{K,\overline{x}} \neq \mathcal{H}_K$. □

4 Covering Number in Learning Theory

In this section we estimate the covering number $\mathcal{N}(I_K(B_R), \eta)$.

Let us first reduce the problem for the subset B_R of an infinitely dimensional space $\mathcal{H}_K$ onto that of a finitely dimensional space.

Lemma 2 *Under the assumption of Theorem 1 and (1.2), for any $x_0 \in X$ and $f \in B_R$, there hold*

$$\big|D^\alpha f(x_0)\big| \leq R\sqrt{C_0}2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2} \quad \forall \alpha \in \mathbb{Z}_+^n \tag{4.1}$$

and

$$\left\| f(x) - \sum_{|\alpha|\leq N} \frac{D^\alpha f(x_0)}{\alpha!}(x - x_0)^\alpha \right\|_{C(U(x_0,\widetilde{r}/2))} \leq \frac{\eta}{2}, \tag{4.2}$$

where $N = N_\eta$ is the integer satisfying

$$\log \frac{5R\sqrt{C_0}}{\eta} / \log\sqrt{2} \leq N < 1 + \log \frac{5R\sqrt{C_0}}{\eta} / \log\sqrt{2}. \tag{4.3}$$

Proof The bound (4.1) follows from Lemma *1* and (1.2).

By Theorem 1, we have

$$f(x)=\sum_{\alpha\in\mathbb{Z}_+^n}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha \quad \forall x\in U\big(x_0,\widetilde{r}\big).$$

This in connection with (4.1) tells us that for $x\in U(x_0,\widetilde{r}/2)\cap X$

$$\begin{aligned}&\left|f(x)-\sum_{|\alpha|\le N}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha\right|\\&\quad\le\sum_{|\alpha|>N}\frac{R2^{|\alpha|}\sqrt{\alpha!C_0|\alpha|!}}{\alpha!}\left(\frac{2}{r}\right)^{|\alpha|/2}\big|(x-x_0)^\alpha\big|.\end{aligned}$$

Since $\widetilde{r}\le\sqrt{r}/(2n)$ and

$$\sum_{|\alpha|=k}\frac{|\alpha|!}{\alpha!}\big|(x-x_0)^\alpha\big|=\left(\sum_{i=1}^n\big|x^i-x_0^i\big|\right)^{|\alpha|}\le\big(\sqrt{n}|x-x_0|\big)^{|\alpha|}, \tag{4.4}$$

the above expression can be bounded by

$$\begin{aligned}&R\sqrt{C_0}\sum_{k>N}2^k\left(\frac{2}{r}\right)^{k/2}\sum_{|\alpha|=k}\frac{|\alpha|!}{\alpha!}\big|(x-x_0)^\alpha\big|\\&\quad\le R\sqrt{C_0}\sum_{k=N+1}^\infty\left(\frac{1}{\sqrt{2}}\right)^k=R\sqrt{C_0}\frac{1}{\sqrt{2}-1}\left(\frac{1}{\sqrt{2}}\right)^N\end{aligned}$$

which is bounded by $\eta/2$ since (4.3) yields $\big(\frac{1}{\sqrt{2}}\big)^N\le\eta/(5R\sqrt{C_0})$. □

By (4.4), the following estimate is trivial.

Lemma 3 *Let* $\widetilde{r}=\min\{\frac{\sqrt{r}}{2n},\sqrt{r+D}-\sqrt{D}\}$ *for some* $r,D>0$. *Then for any sequences* $\{c_\alpha,d_\alpha\}_{|\alpha|\le N}$, *there holds*

$$\begin{aligned}&\left\|\sum_{|\alpha|\le N}\frac{c_\alpha-d_\alpha}{\alpha!}2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}\,(x-x_0)^\alpha\right\|_{C(U(x_0,\widetilde{r}/2))}\\&\quad\le\sum_{k=0}^N\big(1/\sqrt{2}\big)^k\Big\{\max_{|\alpha|\le N}|c_\alpha-d_\alpha|\Big\}\le 4\max_{|\alpha|\le N}|c_\alpha-d_\alpha|.\end{aligned}$$

Now we can prove our second main result.

Proof of Theorem 2 Denote $m=\mathcal{N}(X,\widetilde{r}/2)$ as the covering number of the compact set X in $\mathbb{R}^n$. Then there exists a set $\{x_i\}_{i=1}^m\subset X$ such that $\cup_{i=1}^m U(x_i,\widetilde{r}/2)\supset X$. With this covering of the domain X, we have

$$\mathcal{N}(I_K(B_R),\eta) \le \Pi_{i=1}^m \mathcal{N}(I_K(B_R)|_{U(x_i,\widetilde{r}/2)},\eta).$$

So we need to estimate the covering number $\mathcal{N}(I_K(B_R)|_{U(x_i,\widetilde{r}/2)},\eta)$ for $I_K(B_R)|_{U(x_i,\widetilde{r}/2)}$, a set of functions on the domain $U(x_i,\widetilde{r}/2)$ for each i. To this end, we use (1.2). Let $i \in \{1,\dots,m\}$.

Applying (4.1) with $x_0 = x_i$ in Lemma 2, we know that for each $f \in B_R$ the sequence

$$\left\{c_{f,\alpha} := D^\alpha f(x_0)/\left(2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}\right)\right\}_{\alpha\in\mathbb{Z}_+^n}$$

satisfies

$$D^\alpha f(x_0) = c_{f,\alpha}2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2} \quad \text{and} \quad |c_{f,\alpha}| \le R\sqrt{C_0} \quad \forall \alpha \in \mathbb{Z}_+^n.$$

Let m_η be the integer satisfying

$$\frac{8R\sqrt{C_0}}{\eta} - 1 < m_\eta \le \frac{8R\sqrt{C_0}}{\eta}. \tag{4.5}$$

Let $N = N_\eta$ be given by (4.3). Then for each $f \in B_R$ there exists a sequence

$$\left\{j_{f,\alpha} \in \{-m_\eta, -m_\eta+1,\dots,m_\eta\}\right\}_{\alpha\in\mathbb{Z}_+^n}$$

such that for each $\alpha \in \mathbb{Z}_+^n$,

$$\left|c_{f,\alpha} - \frac{\eta}{8}j_{f,\alpha}\right| \le \frac{\eta}{8}.$$

This in connection with Lemma *3* implies that

$$\left\|\sum_{|\alpha|\le N}\frac{c_{f,\alpha}-\frac{\eta}{8}j_{f,\alpha}}{\alpha!}2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}(x-x_i)^\alpha\right\|_{C(U(x_i,\widetilde{r}/2))} \le \frac{\eta}{2}.$$

That is,

$$\begin{aligned}&\Bigg\|\sum_{|\alpha|\le N}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha \\ &\quad - \sum_{|\alpha|\le N}\frac{\frac{\eta}{8}j_{f,\alpha}}{\alpha!}2^{|\alpha|}\sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}(x-x_i)^\alpha\Bigg\|_{C(U(x_i,\widetilde{r}/2))} \le \frac{\eta}{2}.\end{aligned} \tag{4.6}$$

By (4.2) of Lemma 2, we see that

$$\left\|f(x) - \sum_{|\alpha|\le N}\frac{D^\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha\right\|_{C(U(x_i,\widetilde{r}/2))} \le \frac{\eta}{2}. \tag{4.7}$$

Combining (4.6) and (4.7), we know that for each $f \in B_R$ there exists a sequence

$$\left\{ j_{f,\alpha} \in \{-m_\eta, -m_\eta + 1, \dots, m_\eta\} \right\}_{|\alpha| \le N}$$

such that

$$\left\| f(x) - \sum_{|\alpha| \le N} \frac{\frac{\eta}{8} j_{f,\alpha}}{\alpha!} 2^{|\alpha|} \sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}\,(x - x_i)^\alpha \right\|_{C(U(x_i,\widetilde{r}/2))} \le \eta.$$

Thus,

$$\left\{ \sum_{|\alpha| \le N} \frac{\frac{\eta}{8} j_\alpha}{\alpha!} 2^{|\alpha|} \sqrt{\alpha!|\alpha|!}\,(2/r)^{|\alpha|/2}\,(x - x_i)^\alpha : j_\alpha \in \{-m_\eta, -m_\eta + 1, \dots, m_\eta\} \right\}$$

is an η-net of the set $I_K(B_R)|_{U(x_i,\widetilde{r}/2)}$. Therefore,

$$\mathcal{N}(I_K(B_R), \eta) \le \Pi_{i=1}^m \left(2m_\eta + 1\right)^{\widetilde{N}}$$

where $\widetilde{N}$ is the number of elements in the set $\{\alpha \in \mathbb{Z}_+^n : |\alpha| \le N\}$, which is bounded by $(N+1)^n$. It follows that

$$\begin{aligned} \log \mathcal{N}(I_K(B_R), \eta) &\le m(N+1)^n \log(2m_\eta + 1) \\ &\le \mathcal{N}(X, \widetilde{r}/2) \left(2 + \log \frac{5R\sqrt{C_0}}{\eta} / \log \sqrt{2}\right)^n \log\left(\frac{16R\sqrt{C_0}}{\eta} + 1\right). \end{aligned}$$

Then the desired bound (1.3) follows. This proves Theorem 2. □

References

1. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
2. Belkin, M., Niyogi, P.: Semi-supervised learning on Riemannian manifolds. Mach. Learn. **56**, 209–239 (2004)
3. Cucker, F., Zhou, D.X.: Learning Theory: An Approximation Theory Viewpoint. Cambridge University Press, Cambridge (2007)
4. De Vito, E., Caponnetto, A., Rosasco, L.: Model selection for regularized least-squares algorithm in learning theory. Found. Comput. Math. **5**, 59–85 (2005)
5. Evgeniou, T., Pontil, M., Poggio, T.: Regularization networks and support vector machines. Adv. Comput. Math. **13**, 1–50 (2000)
6. Hardin, D., Tsamardinos, I., Aliferis, C.F.: A theoretical characterization of linear SVM-based feature selection. In: Proc. of the 21st Int. Conf. on Machine Learning, Banff, Canada (2004)
7. Harris, J.: Algebraic Geometry. Springer, New York (1995)
8. Micchelli, C.A.: Interpolation of scattered data: Distance matrices and conditionally positive definite functions. Constr. Approx. **2**, 11–22 (1986)
9. Mukherjee, S., Wu, Q.: Estimation of gradients and coordinate covariances in classification. J. Mach. Learn. Res. **7**, 2481–2514 (2006)
10. Mukherjee, S., Zhou, D.X.: Learning coordinate covariances via gradients. J. Mach. Learn. Res. **7**, 519–549 (2006)
11. Saitoh, S.: Integral Transforms, Reproducing Kernels and their Applications. Longman, Harlow (1997)

12. Schaback, R., Werner, J.: Linearly constrained reconstruction of functions by kernels, with applications to machine learning. Adv. Comput. Math. **25**, 237–258 (2006)
13. Schwartz, L.: Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). J. Anal. Math. **13**, 115–256 (1964)
14. Smale, S., Zhou, D.X.: Estimating the approximation error in learning theory. Anal. Appl. **1**, 17–41 (2003)
15. Smale, S., Zhou, D.X.: Shannon sampling and function reconstruction from point values. Bull. Am. Math. Soc. **41**, 279–305 (2004)
16. Smale, S., Zhou, D.X.: Learning theory estimates via integral operators and their approximations. Constr. Approx. **26**, 153–172 (2007)
17. Steinwart, I., Hush, D., Scovel, C.: An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. IEEE Trans. Inf. Theory **52**, 4635–4643 (2006)
18. Ying, Y., Zhou, D.X.: Online regularized classification algorithms. IEEE Trans. Inf. Theory **52**, 4775–4788 (2006)
19. Zhou, D.X.: Capacity of reproducing kernel spaces in learning theory. IEEE Trans. Inf. Theory **49**, 1743–1752 (2003)
20. Zhou, D.X.: The covering number in learning theory. J. Complex. **18**, 739–767 (2002)
21. Zhou, D.X.: Derivative reproducing properties for kernel methods in learning theory. J. Comput. Appl. Math. (2007). doi:10.1016/j.cam.2007.08.023