

Research Article

The Learning Rates of Regularized Regression Based on Reproducing Kernel Banach Spaces

Baohuai Sheng¹ and Peixin Ye²

¹ Department of Mathematics, Shaoxing University, Shaoxing 312000, China

² School of Mathematics and LPMC, Nankai University, Tianjin 300071, China

Correspondence should be addressed to Baohuai Sheng; bhsheng@usx.edu.cn

Received 13 August 2013; Accepted 16 October 2013

Academic Editor: Qiang Wu

Copyright © 2013 B. Sheng and P. Ye. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the convergence behavior of regularized regression based on reproducing kernel Banach spaces (RKBSs). The convex inequality of uniform convex Banach spaces is used to show the robustness of the optimal solution with respect to the distributions. The learning rates are derived in terms of the covering number and K -functional.

1. Introduction

Recently, there is an increasing research interest in learning with abstract functional spaces, and considerable work has been done in [1–3] and so on.

Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a normed vector space consisting of real functions on a compact distance space $(X, d(\cdot, \cdot))$, and let $M > 0$ be a given positive number. Let $z = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m \in Z^m = (X \times [-M, M])^m$ be a finite set of samples drawn independently and identically (i.i.d.) according to a distribution $\rho(x, y)$ on Z . Then, the regularized learning scheme associating with a given hypothesis space \mathcal{B} and the least square loss is

$$f_z := \arg \min_{f \in \mathcal{B}} \left(\frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 + \frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right), \quad (1)$$

where $q \geq 1$ is a given real number. The unknown Borel probability distribution $\rho(x, y)$ can be decomposed into $\rho(y | x)$ and $\rho_X(x)$, where $\rho(y | x)$ is the conditional probability of ρ at $x \in X$ and $\rho_X(x)$ is the marginal probability on X .

The regression function corresponding to the least square loss is

$$f_\rho(x) = \int_{[-M, M]} y d\rho(y | x), \quad (2)$$

which satisfies

$$f_\rho(x) := \arg \min_f \mathcal{E}_\rho(f) = \arg \min_f \int_Z (y - f(x))^2 d\rho. \quad (3)$$

When the hypothesis spaces \mathcal{B} in (1) are reproducing kernel Banach spaces, we call it the RKBSs based on regularized regression learning defined by [4, 5] recently. The represented theorem related closely to regularized learning is studied in case that \mathcal{B} is an RKBS, and the discussions are extended to the generalized semi-inner-product RKBSs in [6].

In the present paper, we will provide an investigation on the learning rates of scheme (1) when \mathcal{B} is an RKBS with uniform convexity. The paper is organized as follows. In Section 2, we show the main results of the present paper. The robustness is studied in Section 3, and the sample errors are bounded in Section 4. The approximation error boils down to a K -functional. The learning rates are bounded in Section 5.

For a given real number $p \geq 1$, we denote by $L^p(\rho_X)$ the class of ρ_X -measurable functions f satisfying $\|f\|_{L^p(\rho_X)} = (\int_X |f(x)|^p d\rho_X) < +\infty$.

We say $A = O(B)$ if there is a constant $C > 0$ such that $A/B \leq C$. We say $A \sim B$ if both $A = O(B)$ and $B = O(A)$.

2. Notions and Results

To state the results of the present paper, we first introduce some notions as follows.

2.1. The RKBSs. We denote by \mathcal{B} the Banach space with dual space \mathcal{B}^* and norm $\|\cdot\|_{\mathcal{B}}$. For $f \in \mathcal{B}$ and $f^* \in \mathcal{B}^*$, we write $\langle f, f^* \rangle_{\mathcal{B}} = f^*(f)$.

A reproducing kernel Banach space (RKBS) on X is a reflexive Banach space of real functions on X whose dual space \mathcal{B}^* is isometric to a Banach space $\mathcal{B}^\#$ of functions on X , and the point evaluations are continuous functions on both \mathcal{B} and $\mathcal{B}^\#$. It was shown by Theorem 2 of [4] that if B is an RKBS on X , then, there exists uniquely a function $K : X \times X \rightarrow \mathfrak{R}$ called the reproducing kernel of \mathcal{B} satisfying the following:

- (i) $K(x, \cdot) \in \mathcal{B}$, $K(\cdot, x) \in \mathcal{B}^*$, $x \in X$;
- (ii) $f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{B}}$, $f^*(x) = \langle K(x, \cdot), f^* \rangle_{\mathcal{B}}$, $f^* \in \mathcal{B}^*$, $x \in X$.
- (iii) The linear span of $\{K(x, \cdot) : x \in X\}$ is dense in \mathcal{B} , namely,

$$\overline{\text{span}} \{K(x, \cdot) : x \in X\} = \mathcal{B}. \quad (4)$$

- (iv) The linear span of $\{K(\cdot, x) : x \in X\}$ is dense in \mathcal{B}^* , namely,

$$\overline{\text{span}} \{K(\cdot, x) : x \in X\} = \mathcal{B}^*. \quad (5)$$

- (v) For all $x, y \in X$ there holds $K(x, y) = \langle K(x, \cdot), K(\cdot, y) \rangle_{\mathcal{B}}$.

When \mathcal{B} is an RKHS, K is indeed the reproducing kernel in the usual sense (see [7]).

Since \mathcal{B} is a reflective Banach space, we have

$$\langle f, f^* \rangle_{\mathcal{B}} \leq \|f\|_{\mathcal{B}} \times \|f^*\|_{\mathcal{B}^*}, \quad f \in \mathcal{B}, \quad f^* \in \mathcal{B}^*, \quad (6)$$

$$\|f^*\|_{\mathcal{B}^*} = \sup_{\|f\|_{\mathcal{B}} \leq 1} \langle f, f^* \rangle_{\mathcal{B}}. \quad (7)$$

A way of producing reproducing kernel spaces in L^p spaces by the idempotent integral operators was provided in [8]. In the present paper, we provide a method to construct RKBSs by orthogonal function series.

Example 1. Let $X = [a, b]$ be a given closed interval and let, $\{\varphi_k(x)\}_{k=0}^{+\infty}$ be a sequence of continuous functions on $[a, b]$ satisfying the following:

- (i) $\varphi_k \in L^p(\rho)$ for $k = 0, 1, 2, \dots$;
- (ii) φ_k and φ_l are orthonormal (in $L^2(\rho)$) when $l \neq k$;
- (iii) $\text{span}(\{\varphi_k(x)\}_{k=0}^{+\infty})$ is dense in $L^p(\rho)$ for $1 < p < +\infty$.

Let $\{\lambda_l\}_{l=0}^{+\infty}$ be a given positive real number sequence satisfying $\sum_{l=0}^{+\infty} \lambda_l \leq 1$. Define

$$K(x, y) = \sum_{l=0}^{\infty} \lambda_l \varphi_l(x) \varphi_l(y), \quad x, y \in X, \quad (8)$$

and the functional class $\mathcal{B}_K^{(p)}$ on X by

$$\mathcal{B}_K^{(p)} = \left\{ f \in C(X) : \|f\|_{\mathcal{B}_K^{(p)}} = \left(\sum_{l=0}^{\infty} \frac{|a_l(f)|^{p'}}{\lambda_l} \right)^{1/p'} < +\infty \right\}, \quad (9)$$

where $a_k(f) = \int_{[a,b]} f(y) \varphi_k(y) d\rho(y)$, $k \in N$. We define the space $\mathcal{B}_K^{(p')}$ for $p' = p/(p-1)$ in an analogous way.

We have the following proposition.

Proposition 2. Define a bivariate operation on $\mathcal{B}_K^{(p)}$ and $\mathcal{B}_K^{(p')}$ by

$$\langle f, g \rangle = \sum_{l=0}^{\infty} \frac{a_l(f) a_l(g)}{\lambda_l}, \quad f \in \mathcal{B}_K^{(p)}, \quad g \in \mathcal{B}_K^{(p')}. \quad (10)$$

Then, $\mathcal{B}_K^{(p)}$ is a reproducing kernel Banach space with reproducing kernel $K(x, y)$.

Proof. Let $l_K^p = \{a = \{a_k\}_{k=0}^{+\infty} : \|a\|_{l_K^p} = (\sum_{l=0}^{\infty} (|a_l|^p / \lambda_l))^{1/p} < +\infty\}$ and $l_K^{p'}$ be defined in an analogous way. Then, both l_K^p and $l_K^{p'}$ are Banach spaces and $(l_K^p)^* = l_K^{p'}$ and $(l_K^{p'})^* = l_K^p$. \square

By (9) we know $\mathcal{B}_K^{(p)}$ and $l_K^{p'}$ are isometric isomorphisms. Therefore, $\mathcal{B}_K^{(p)}$ are Banach spaces.

Since $a_l(K(\cdot, y)) = \lambda_l \varphi_l(y)$, we have for $f \in \mathcal{B}_K^{(p)}$ that

$$\begin{aligned} \langle f, K(\cdot, y) \rangle &= \sum_{l=0}^{\infty} \frac{a_l(f) a_l(K(\cdot, y))}{\lambda_l} \\ &= \sum_{l=0}^{\infty} a_l(f) \varphi_l(y) = f(y). \end{aligned} \quad (11)$$

By the same way, we have for any $g \in \mathcal{B}_K^{(p')}$ that $\langle K(x, \cdot), g \rangle = g(x)$; that is, the reproducing property holds.

2.2. The Uniform Convexity. In this subsection, we focus on some notions in convex analysis and Banach geometry theory.

Let $F(f) : \mathcal{B} \rightarrow \mathfrak{R}$ be a convex function. Then,

$$\begin{aligned} \partial F(f) &= \{ \xi \in \mathcal{B}^* : F(f') - F(f) \\ &\geq \langle f' - f, \xi \rangle_{\mathcal{B}}, \quad \forall f' \in \mathcal{B} \} \neq \emptyset. \end{aligned} \quad (12)$$

We call $\partial F(f)$ the subdifferential of $F(f)$ at $f \in \mathcal{B}$. If $\xi \in \partial F(f)$, then, we call ξ a subgradient of F at f .

A well-known result is that f_0 is a minimal value point of a convex function $F(f)$ on \mathcal{B} if and only if $0 \in \partial F(f_0)$ (see [9]).

A Banach space \mathcal{B} is called q -uniform convex if there are constants $q > 0, c > 0$ such that the modulus defined by

$$\begin{aligned} \delta_{\mathcal{B}}(\varepsilon) &= \inf \left(1 - \frac{\|f + g\|_{\mathcal{B}}}{2} : \|f\|_{\mathcal{B}} = \|g\|_{\mathcal{B}} = 1, \|f - g\|_{\mathcal{B}} = \varepsilon \right), \\ &\varepsilon \in (0, 2] \end{aligned} \tag{13}$$

satisfies $\delta_{\mathcal{B}}(\varepsilon) \geq c\varepsilon^q$. In particular, any Hilbert spaces are 2-uniform convex Banach spaces.

Define $J_q(f)_{\mathcal{B}} = \partial((1/q)\|f\|_{\mathcal{B}}^q)$. Then, by (28) in Corollary 1 of [10] we know \mathcal{B} is q -uniform convex if and only if there is a positive constant $c_q > 0$ such that for all $f, g \in \mathcal{B}$ and all $j_q(f) \in J_q(f)_{\mathcal{B}}$ there holds

$$\begin{aligned} \frac{1}{q} (\|f\|_{\mathcal{B}}^q - \|g\|_{\mathcal{B}}^q) &\geq \langle j_q(g), f - g \rangle_{\mathcal{B}} + c_q \|f - g\|_{\mathcal{B}}^q, \\ &\forall f, g \in \mathcal{B}. \end{aligned} \tag{14}$$

In [11–14] we know that, for a given $1 < p < +\infty$, the space l_p , the Lebesgue spaces L_p and the Sobolev space W_p^m are $\max\{2, p\}$ -uniform convex. Also, let $\mathcal{B}_K^{(p)}$ and $\mathcal{B}_K^{(p')}$ be defined as in Section 2.1. Then, by the fact that $\mathcal{B}_K^{(p)}$ and $l_p^{p'}$ are isometric isomorphisms, we know $\mathcal{B}_K^{(p)}$ is 2-uniform convex if $p > 2$ and p' -uniform convex if $1 < p \leq 2$. Therefore, we know $\mathcal{B}_K^{(p)}$ is a q -uniform convex Banach space, where q is 2 if $p > 2$ and its value is $p/(p - 1)$ if $1 < p \leq 2$.

2.3. Main Results. Let S be a distance space and $\eta > 0$. The covering number $\mathcal{N}(S, \eta)$ is defined to be the minimal positive integer number l such that there exists l disk in S with radius η covering S .

We say a compact subset E in a distance space $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ has logarithmic complexity exponent $s \geq 0$ if there is a constant $c_s > 0$ such that the closed ball of radius R centered at origin, that is, $\mathcal{B}_R = \{f \in E : \|f\|_{\mathcal{B}} \leq R\}$, satisfies

$$\log \mathcal{N}(\mathcal{B}_R, \eta) \leq c_s \left(\frac{R}{\eta}\right)^s, \quad \forall \eta > 0. \tag{15}$$

Now we are in a position to present the main results of this paper.

Theorem 3. *Let \mathcal{B} be an RKBS with q -uniform convexity and a reproducing kernel $K(\cdot, x)$ which is uniform continuous on X in terms of the norm $\|\cdot\|_{\mathcal{B}^*}$, that is, $\|K(\cdot, x)\|_{\mathcal{B}^*} \cdot \|K(\cdot, x)\|_{\mathcal{B}^*}$ is a uniform continuous function on X , and there is a constant $k > 0$ such that $\|K(\cdot, x)\|_{\mathcal{B}^*} \leq k$ holds for all $x \in X$. Let f_z*

be the unique minimizer of scheme (1). If $f_\rho \in L^2(\rho_X)$, then for any $\varepsilon > 0$ there holds

$$\begin{aligned} \text{Prob}_{z \in Z^m} \left(\|f_z - f_\rho\|_{L^2(\rho_X)} \leq k\varepsilon + \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)} \right) \\ \geq 1 - 2\mathcal{N} \left(\mathcal{B}_1, \frac{\lambda c_q \varepsilon^{q-1}}{8 \gamma_q} \right) \\ \times \exp \left(-\frac{m\lambda^2 c_q^2 \varepsilon^{2(q-1)}}{32 k^2 \gamma_q^2} \right), \end{aligned} \tag{16}$$

where

$$D_q(f_\rho, \lambda) = \inf_{f \in \mathcal{B}} \left(\|f - f_\rho\|_{L^2(\rho_X)}^2 + \lambda \|f\|_{\mathcal{B}}^q \right) \tag{17}$$

is a K -functional, $\gamma_q := M + k(q D_q(f_\rho, \lambda/q)/\lambda)^{1/q}$ and

$$\mathcal{B}_1 = \{f(x) : f \in \mathcal{B}, \|f\|_{\mathcal{B}} \leq 1\}. \tag{18}$$

The covering number involved in (16) has been studied widely (see [15–19]). In this paper, we assume $\mathcal{N}(\mathcal{B}_R, \eta)$ has the logarithmic complexity.

Theorem 4. *Under the conditions of Theorem 3, if $f_\rho \in L^2(\rho_X)$ and $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ has logarithmic complexity with exponent $s \geq 0$, then for any $\delta \in (0, 1)$, with confidence $1 - \delta$, there holds*

$$\begin{aligned} \|f_z - f_\rho\|_{L^2(\rho_X)} \\ \leq k \left[\left(\frac{64\gamma_q^2 \log(2/\delta)}{m\lambda^2 c_q^2} \right)^{1/2(q-1)} \right. \\ \left. + \left(\frac{8^{2+s} \gamma_q^{2+s} c_s}{m\lambda^{2+s} c_q^{2+s}} \right)^{1/(2+s)(q-1)} \right] \\ + \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)}, \end{aligned} \tag{19}$$

where c_s is defined in (15).

We now give some remarks on Theorems 3 and 4.

(i) In Theorem 3, we require that the kernel $K(x, y)$ is uniform continuous and uniform bounded on X . In fact, a large class of real bivariate functions satisfies these conditions. For example, if the function sequence $\{\varphi_l(x)\}_{l=0}^{+\infty}$ defined in Example 1 is uniformly bounded, that is, $|\varphi_l(x)| \leq 1$ holds for all l and all $x \in [a, b]$, then, kernel $K(x, y)$ is continuous on $[a, b] \times [a, b]$ which turns out that $K(x, y)$ is uniform continuous on $[a, b] \times [a, b]$. Therefore, $|K(x, y)| \leq \sum_{l=0}^{+\infty} \lambda_l \leq 1$ shows that $K(x, y)$ is uniform continuous and bounded with norm $\|\cdot\|_{L^p(\rho)}$.

(ii) By the definition of γ_q , we know that if $D_q(f_\rho, \lambda) = O(\lambda^\beta)$, $0 < \beta \leq 1$, then, $\gamma_q = M + O(q\lambda^{\beta-1})^{1/q}$. It is bounded if $\beta = 1$.

(iii) If \mathcal{B} is a reproducing kernel Hilbert space, then, $q = 2$, $c_q = 1$. Moreover, if $D_q(f_\rho, \lambda) = O(\lambda^\beta)$, $0 < \beta \leq 1$, then, we have by (19) that

$$\begin{aligned} & \|f_z - f_\rho\|_{L^2(\rho_X)} \\ &= O \left[\left(\frac{\log(2/\delta)}{m\lambda^{3-\beta}} \right)^{1/2} + \left(\frac{1}{m\lambda^{(2+s)(3-\beta)/2}} \right)^{1/(2+s)} + \lambda^{\beta/2} \right]. \end{aligned} \quad (20)$$

(iv) We can show a way of bounding the decay rates of $D_q(f_\rho, \lambda)$ for $1 < p \leq 2$. Let $f \in L^p(\rho)$. Then, we have the following Fourier expansion:

$$f(x) \sim \sum_{l=0}^{\infty} a_l(f) \varphi_l(x). \quad (21)$$

Define an operator sequence by

$$V_n(f, x) = \sum_{l=0}^n \lambda_l a_l(f) \varphi_l(x), \quad x \in X. \quad (22)$$

Then, for a given positive integer n we have $a_l(V_n(f)) = \lambda_l a_l(f)$ and

$$\begin{aligned} \|V_n(f)\|_{\mathcal{B}_K^{(p)}} &= \left(\sum_{l=0}^{\infty} \frac{|a_l(V_n(f))|^{p'}}{\lambda_l} \right)^{1/p'} \\ &= \left(\sum_{l=0}^n \lambda_l^{p'-1} |a_l(f)|^{p'} \right)^{1/p'} \\ &\leq \left(\max_{0 \leq l \leq n} \lambda_l^{1-(1/p')} \right) \left(\sum_{l=0}^{+\infty} |a_l(f)|^{p'} \right)^{1/p'} \\ &\leq \left(\max_{0 \leq l \leq n} \lambda_l^{1-(1/p')} \right) \|f\|_{L^p(\rho)}, \end{aligned} \quad (23)$$

where we have used the generalized Bessel inequality (see [20]):

$$\left(\sum_{l=0}^{+\infty} |a_l(f)|^{p'} \right)^{1/p'} \leq \|f\|_{L^p(\rho)}. \quad (24)$$

Also,

$$\begin{aligned} \langle V_n(f), K(\cdot, y) \rangle &= \sum_{l=0}^{\infty} \frac{a_l(V_n(f)) a_l(K(\cdot, y))}{\lambda_l} \\ &= \sum_{l=0}^n \lambda_l a_l(f) \varphi_l(y) = V_n(f, y). \end{aligned} \quad (25)$$

By (25) and (23) we know $V_n(f) \in \mathcal{B}_K^{(p)}$ holds for all positive integers n and, in this case,

$$\begin{aligned} D_q(f_\rho, \lambda) &\leq \|V_n(f_\rho) - f_\rho\|_{L^2(\rho_X)}^2 + \lambda \|V_n(f_\rho)\|_{\mathcal{B}_K^{(p)}}^q \\ &\leq \|V_n(f_\rho) - f_\rho\|_{L^2(\rho_X)}^2 \\ &\quad + \lambda \left(\max_{0 \leq l \leq n} \lambda_l^{1-(1/p')} \right)^q \|f\|_{L^p(\rho)}. \end{aligned} \quad (26)$$

One can choose suitable n such that it depends upon the sample number m and obtain the decay rates when $m \rightarrow +\infty$. There are many choices for the type of operator (22). For example, the Bernstein-Durrmeyer operators (see, e.g., [21–23]) and the de la Vallée-Poussin sum operators are such types (see [24]). This method was first provided by [25] and was extended in [26, 27].

(v) We know from [19] that the RKHSs with logarithmic complexity with exponent $s \geq 0$ exist. By Corollary 4.1 and Theorem 2.1 of [16] we know that if λ_l satisfy $\lambda_l \sim 1/(1+l)^\alpha$, $\alpha > 1$, then, the covering number of $\mathcal{B}_K^{(2)}$ may attain the decay of complexity exponent. In a recent paper (see [28]), Guntuboyina and Sen showed that the set of all convex functions defined on $[a, b]^d$ that are uniform bounded has the logarithmic complexity exponent $d/2$ in the L_p -metric.

3. Robustness

Robustness is a quantitative description of the solutions on the distributions.

Define the ρ -control integral regularized model corresponding to (1) by

$$f^{(\rho)} := \arg \min_{f \in \mathcal{B}} \left(\mathcal{E}_\rho(f) + \frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right), \quad (27)$$

where $\mathcal{E}_\rho(f)$ is defined in (3). Then, $f^{(\rho)}$ is influenced by the distributions ρ . For any bounded ρ -measurable function $f(x, y)$ on Z , we define the empirical measure $\gamma_z(x, y)$ as follows:

$$E_{\gamma_z} [f(x, y)] = \int_Z f(x, y) d\gamma_z = \frac{1}{m} \sum_{i=1}^m f(x_i, y_i). \quad (28)$$

Then, $f_z = f^{(\gamma_z)}$. We give the following theorem.

Theorem 5. Let \mathcal{B} be an RKBS with q -uniform convexity and the reproducing kernel $K(x, y)$, and let $f^{(\rho)}$ and $f^{(\gamma)}$ be the

solutions of scheme (27) with respect to distributions ρ and γ , respectively. Then,

$$\begin{aligned} & \|f^{(\rho)} - f^{(\gamma)}\|_{\mathcal{B}} \\ & \leq \left(\frac{2}{\lambda c_q} \times \left\| \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \right. \\ & \quad \left. \left. - \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\|_{\mathcal{B}^*} \right)^{1/(q-1)}, \end{aligned} \tag{29}$$

where c_q is the constant defined in (14).

Theorem 5 shows how ρ influences the unique solution $f^{(\rho)}$.

To prove Theorem 5, we need the following lemmas.

Lemma 6. Under the conditions of Theorem 5, there holds

$$\partial \mathcal{E}_\rho(f_0) = \left\{ -2 \int_Z (y - f_0(x)) K(\cdot, x) d\rho \right\}, \quad f_0 \in \mathcal{B}, \tag{30}$$

where the point \cdot in $K(\cdot, x)$ means $K(\cdot, x) \in \mathcal{B}$ for any $x \in X$.

Proof. We restate the following statement.

Let $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ be a Banach space, $F(f) : \mathcal{B} \rightarrow \mathfrak{R} \cup \{\mp\infty\}$ be a real function. We say F is Gateaux differentiable at $f_0 \in \mathcal{B}$ if there is an $\xi \in \mathcal{B}^*$ such that for any $g \in \mathcal{B}$ there holds

$$\lim_{t \rightarrow 0} \frac{F(f_0 + tg) - F(f_0)}{t} = \langle g, \xi \rangle_{\mathcal{B}}, \tag{31}$$

and write $F'_G(f_0) = \xi$. By [29] we know that if F is convex on \mathcal{B} and is Gateaux differentiable at $f_0 \in \mathcal{B}$, then, $\partial F(f_0) = \{F'_G(f_0)\}$.

By equality

$$x^2 - y^2 = 2y(x - y) + (x - y)^2, \quad x, y \in \mathfrak{R}, \tag{32}$$

we have for any $g(x) = \langle g, K(\cdot, x) \rangle_{\mathcal{B}} \in \mathcal{B}$ that

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{\mathcal{E}_\rho(f_0 + tg) - \mathcal{E}_\rho(f_0)}{t} \\ & = \lim_{t \rightarrow 0} \frac{\int_Z (y - f_0(x) - tg(x))^2 d\rho - \int_Z (y - f_0(x))^2 d\rho}{t} \\ & = \lim_{t \rightarrow 0} \frac{-2t \int_Z (y - f_0(x)) g(x) d\rho + t^2 \int_Z g(x)^2 d\rho}{t} \\ & = -2 \int_Z (y - f_0(x)) g(x) d\rho \\ & = \left\langle g, -2 \int_Z (y - f_0(x)) K(\cdot, x) d\rho \right\rangle_{\mathcal{B}}. \end{aligned} \tag{33}$$

Since $\mathcal{E}_\rho(f)$ is a convex function on \mathcal{B} , we know (30) holds. \square

Lemma 7. Take $J_q(f) = \partial((1/q)\|f\|_{\mathcal{B}}^q)$. Then, under the conditions of Theorem 5, there hold the following.

(i) There exists uniquely a minimizer $f^{(\rho)}$ of the problem (27) and

$$\|f^{(\rho)}\|_{\mathcal{B}} \leq \left(\frac{qD_q(f_\rho, \lambda/q)}{\lambda} \right)^{1/q}. \tag{34}$$

(ii) There is a $j_q(f^{(\rho)}) \in J_q(f^{(\rho)})$ such that

$$\lambda j_q(f^{(\rho)}) = 2 \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho. \tag{35}$$

Proof. The uniqueness of the minimizer can be obtained by the fact that (27) is a strict convex optimization problem. By the definition of $f^{(\rho)}$, we have

$$\begin{aligned} \frac{\lambda}{q} \|f^{(\rho)}\|_{\mathcal{B}}^q & \leq \frac{\lambda}{q} \|f^{(\rho)}\|_{\mathcal{B}}^q + \mathcal{E}_\rho(f^{(\rho)}) - \mathcal{E}_\rho(f_\rho) \\ & = D_q\left(f_\rho, \frac{\lambda}{q}\right). \end{aligned} \tag{36}$$

We then have (34). \square

Proof of (35). Since $f^{(\rho)}$ is the unique solution of (27), we have

$$0 \in \partial_f \left(\mathcal{E}_\rho(f) + \frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right) \Big|_{f=f^{(\rho)}}. \tag{37}$$

Notice that both $\mathcal{E}_\rho(f)$ and $\|f\|_{\mathcal{B}}^q$ are convex functions about f on \mathcal{B} . We have

$$\begin{aligned} & \partial_f \left(\mathcal{E}_\rho(f) + \frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right) \Big|_{f=f^{(\rho)}} \\ & = \partial_f(\mathcal{E}_\rho(f)) \Big|_{f=f^{(\rho)}} + \partial_f \left(\frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right) \Big|_{f=f^{(\rho)}}. \end{aligned} \tag{38}$$

By (30), we know that (37) leads to

$$0 \in \left\{ -2 \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right\} + \lambda J_q(f^{(\rho)}). \tag{39}$$

Therefore, there is $j_q(f^{(\rho)}) \in J_q(f^{(\rho)})$ such that (35) holds. \square

Lemma 8. Let \mathcal{B} be an RKBS satisfying the conditions of Theorem 3. Then,

$$|f(x)| \leq k \|f\|_{\mathcal{B}}, \quad x \in X, f \in \mathcal{B}. \tag{40}$$

Proof. The reproducing property and (16) give

$$|f(x)| = |\langle f, K(\cdot, x) \rangle_{\mathcal{B}}| \leq \|f\|_{\mathcal{B}} \times \|K(\cdot, x)\|_{\mathcal{B}^*}. \tag{41}$$

Then, the fact $\|K(\cdot, x)\|_{\mathcal{B}^*} \leq k$ gives (40). \square

Lemma 9. Let $K(x, y)$ be the reproducing kernel of \mathcal{B} , and $K(\cdot, x)$ is uniform continuous about x on X in norm $\|\cdot\|_{\mathcal{B}^*}$, $R > 0$ be a given real number. Then, the ball $\mathcal{B}_R = \{f \in \mathcal{B} : \|f\|_{\mathcal{B}} \leq R\}$ is a compact subset of $C(X)$.

Proof. Since X is a compact distance space, so is $X \times X$. Since $K(\cdot, x)$ is uniform continuous about x in norm $\|\cdot\|_{\mathcal{B}^*}$, we know that for any $\epsilon > 0$ there is a $\delta > 0$ such that for all $x, x' \in X$ with $d(x, x') < \delta$, we have

$$\|K(\cdot, x) - K(\cdot, x')\|_{\mathcal{B}^*} < \epsilon \quad (42)$$

and for any $f \in \mathcal{B}_R$ holds

$$\begin{aligned} & |f(x) - f(x')| \\ &= |\langle f, K(\cdot, x) \rangle_{\mathcal{B}} - \langle f, K(\cdot, x') \rangle_{\mathcal{B}}| \\ &= |\langle f, K(\cdot, x) - K(\cdot, x') \rangle_{\mathcal{B}}| \\ &\leq \|f\|_{\mathcal{B}} \times \|K(\cdot, x) - K(\cdot, x')\|_{\mathcal{B}^*} \leq R\epsilon. \end{aligned} \quad (43)$$

By (43), we know that \mathcal{B}_R is a closed, bounded, and equicontinuous set. Therefore, \mathcal{B}_R is a compact set of $C(X)$. \square

Proof of Theorem 5. By the definition of $\partial \mathcal{E}_\gamma(f^{(\rho)})$ and (30) we know

$$\begin{aligned} & \mathcal{E}_\gamma(f^{(\gamma)}) - \mathcal{E}_\gamma(f^{(\rho)}) \\ &\geq \left\langle f^{(\gamma)} - f^{(\rho)}, -2 \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\rangle_{\mathcal{B}}. \end{aligned} \quad (44)$$

Also, by (44) and the definitions of $f^{(\rho)}$ and $f^{(\gamma)}$ we have

$$\begin{aligned} 0 &\geq \left(\int_Z (y - f^{(\gamma)}(x))^2 d\gamma + \frac{\lambda}{q} \|f^{(\gamma)}\|_{\mathcal{B}}^q \right) \\ &\quad - \left(\int_Z (y - f^{(\rho)}(x))^2 d\gamma + \frac{\lambda}{q} \|f^{(\rho)}\|_{\mathcal{B}}^q \right) \\ &= \mathcal{E}_\gamma(f^{(\gamma)}) - \mathcal{E}_\gamma(f^{(\rho)}) + \frac{\lambda}{q} (\|f^{(\gamma)}\|_{\mathcal{B}}^q - \|f^{(\rho)}\|_{\mathcal{B}}^q) \\ &\geq 2 \left\langle f^{(\gamma)} - f^{(\rho)}, - \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\rangle_{\mathcal{B}} \\ &\quad + \frac{\lambda}{q} (\|f^{(\gamma)}\|_{\mathcal{B}}^q - \|f^{(\rho)}\|_{\mathcal{B}}^q). \end{aligned} \quad (45)$$

Since \mathcal{B} is q -uniform convex, we have by (14) and the definition of $j_q(f^{(\rho)})$ that

$$\begin{aligned} \frac{1}{q} (\|f^{(\gamma)}\|_{\mathcal{B}}^q - \|f^{(\rho)}\|_{\mathcal{B}}^q) &\geq \langle f^{(\gamma)} - f^{(\rho)}, j_q(f^{(\rho)}) \rangle_{\mathcal{B}} \\ &\quad + c_q \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}}^q. \end{aligned} \quad (46)$$

Combining (46) with (45), we have

$$\begin{aligned} 0 &\geq \left(\int_Z (y - f^{(\gamma)}(x))^2 d\gamma + \frac{\lambda}{q} \|f^{(\gamma)}\|_{\mathcal{B}}^q \right) \\ &\quad - \left(\int_Z (y - f^{(\rho)}(x))^2 d\gamma + \frac{\lambda}{q} \|f^{(\rho)}\|_{\mathcal{B}}^q \right) \\ &\geq 2 \left\langle f^{(\rho)} - f^{(\gamma)}, \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\rangle_{\mathcal{B}} \\ &\quad + \lambda \langle f^{(\gamma)} - f^{(\rho)}, j_q(f^{(\rho)}) \rangle_{\mathcal{B}} \\ &\quad + \lambda c_q \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}}^q \\ &\stackrel{(35)}{=} 2 \left\langle f^{(\rho)} - f^{(\gamma)}, \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\rangle_{\mathcal{B}} \\ &\quad - 2 \left\langle f^{(\rho)} - f^{(\gamma)}, \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right\rangle_{\mathcal{B}} \\ &\quad + \lambda c_q \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}}^q \\ &= 2 \left\langle \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right. \\ &\quad \left. - \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho, f^{(\rho)} - f^{(\gamma)} \right\rangle_{\mathcal{B}} \\ &\quad + \lambda c_q \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}}^q. \end{aligned} \quad (47)$$

It follows that

$$\begin{aligned} & \lambda c_q \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}}^q \\ &\leq 2 \left\langle f^{(\gamma)} - f^{(\rho)}, \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right. \\ &\quad \left. - \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right\rangle_{\mathcal{B}} \\ &\leq 2 \|f^{(\gamma)} - f^{(\rho)}\|_{\mathcal{B}} \\ &\quad \times \left\| \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \\ &\quad \left. - \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\gamma \right\|_{\mathcal{B}^*}. \end{aligned} \quad (48)$$

We then have (29). \square

4. Sample Error

We give the following sample error bounds.

Theorem 10. Let \mathcal{B} be an RKBS satisfying the conditions of Theorem 3. $f^{(\rho)}$ is the solution of scheme (27) with respect to ρ and f_z is the solution of (1). Then, for all $\epsilon > 0$ there hold

$$\begin{aligned} & \text{Prob}_{z \in Z^m} \left\{ \|f^{(\rho)} - f_z\|_{\mathcal{B}} \leq \epsilon \right\} \\ & \geq 1 - 2\mathcal{N} \left(\mathcal{F}, \frac{\lambda c_q \epsilon^{q-1}}{8} \right) \exp \left(-\frac{m \lambda^2 c_q^2 \epsilon^{2(q-1)}}{32 \gamma_q^2} \right), \end{aligned} \quad (49)$$

where

$$\mathcal{F} = \left\{ h(z) f(x) : f \in \mathcal{B}, \|f\|_{\mathcal{B}} \leq 1, \|h\|_{C(Z)} \leq \gamma_q \right\}. \quad (50)$$

To show Theorem 10, we first give a lemma.

Lemma 11 (see [15]). Let \mathcal{F} be a family of functions from a probability space Z to \mathfrak{R} and $d(\cdot, \cdot)$ a distance on \mathcal{F} . Let $\mathcal{U} \subset Z$ be of full measure and constants $B, L > 0$ such that

- (i) $|\xi(z)| \leq B$ for all $\xi \in \mathcal{F}$ and all $z \in \mathcal{U}$,
- (ii) $|L_z(\xi_1) - L_z(\xi_2)| \leq Ld(\xi_1, \xi_2)$ for all $\xi_1, \xi_2 \in \mathcal{F}$ and all $z \in \mathcal{U}^m$, where

$$L_z(\xi) = \int_Z \xi(z) - \frac{1}{m} \sum_{i=1}^m \xi(z_i). \quad (51)$$

Then, for all $\epsilon > 0$,

$$\begin{aligned} & \text{Prob}_{z \in Z^m} \left\{ \sup_{\xi \in \mathcal{F}} |L_z(\xi)| \leq \epsilon \right\} \geq 1 - \mathcal{N} \left(\mathcal{F}, \frac{\epsilon}{2L} \right) \\ & \times 2 \exp \left(-\frac{m\epsilon^2}{8B^2} \right). \end{aligned} \quad (52)$$

Proof of Theorem 10. Take $\gamma = \gamma_z$ into (29). Then,

$$\begin{aligned} & \|f^{(\rho)} - f_z\|_{\mathcal{B}} \\ & \leq \left(\frac{2}{\lambda c_q} \times \left\| \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \right. \\ & \quad \left. \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) K(\cdot, x_i) \right\|_{\mathcal{B}^*} \right)^{1/(q-1)}. \end{aligned} \quad (53)$$

By (7) and the reproducing property, we have

$$\begin{aligned} & \left\| \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \\ & \quad \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) K(\cdot, x_i) \right\|_{\mathcal{B}^*} \\ & = \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \left\langle f, \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \right. \\ & \quad \left. \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) K(\cdot, x_i) \right\rangle_{\mathcal{B}} \right| \\ & = \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \int_Z (y - f^{(\rho)}(x)) \langle f, K(\cdot, x) \rangle_{\mathcal{B}} d\rho \right. \\ & \quad \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) \langle f, K(\cdot, x_i) \rangle_{\mathcal{B}} \right| \\ & = \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \int_Z (y - f^{(\rho)}(x)) f(x) d\rho \right. \\ & \quad \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) f(x_i) \right|. \end{aligned} \quad (54)$$

Since

$$\begin{aligned} & |(y - f^{(\rho)}(x))| \leq |y| + |f^{(\rho)}(x)| \\ & \leq M + k \left(\frac{qD_q(f_\rho, \lambda/q)}{\lambda} \right)^{1/q} \end{aligned} \quad (55)$$

and (40), we have

$$\begin{aligned} & \left\| \int_Z (y - f^{(\rho)}(x)) K(\cdot, x) d\rho \right. \\ & \quad \left. - \frac{1}{m} \sum_{i=1}^m (y_i - f^{(\rho)}(x_i)) K(\cdot, x_i) \right\|_{\mathcal{B}^*} \\ & \leq \sup_{f \in \mathcal{F}} \left| \int_Z f(z) d\rho - \frac{1}{m} \sum_{i=1}^m f(z_i) \right|. \end{aligned} \quad (56)$$

Define

$$L_z(f) = \int_Z f(z) d\rho - \frac{1}{m} \sum_{i=1}^m f(z_i). \quad (57)$$

Then,

$$\begin{aligned}
& |L_z(f_1) - L_z(f_2)| \\
& \leq \left| \int_Z (f_1(z) - f_2(z)) d\rho \right| \\
& \quad + \frac{1}{m} \sum_{i=1}^m |f_1(z_i) - f_2(z_i)| \\
& \leq 2\|f_1 - f_2\|_{C(Z)}.
\end{aligned} \tag{58}$$

By (52), we have for all $\epsilon > 0$ that

$$\text{Prob}_{z \in Z^m} \left\{ \sup_{f \in \mathcal{F}} |L_z(f)| \geq \epsilon \right\} \leq 2\mathcal{N} \left(\mathcal{F}, \frac{\epsilon}{4} \right) \exp \left(-\frac{m\epsilon^2}{8\gamma_q^2} \right). \tag{59}$$

By (53), (56), and (59), we know

$$\frac{\lambda c_q}{2} \|f^{(\rho)} - f_z\|_{\mathcal{B}}^{q-1} \leq \sup_{f \in \mathcal{F}} \left| \int_Z f(z) d\rho - \frac{1}{m} \sum_{i=1}^m f(z_i) \right|, \tag{60}$$

which gives

$$\begin{aligned}
& \left\{ z \in Z^m : \frac{\lambda c_q}{2} \|f^{(\rho)} - f_z\|_{\mathcal{B}}^{q-1} \geq \epsilon \right\} \\
& \subset \left\{ z \in Z^m : \sup_{f \in \mathcal{F}} |L_z(f)| \geq \epsilon \right\}.
\end{aligned} \tag{61}$$

It follows that

$$\begin{aligned}
& \text{Prob}_{z \in Z^m} \left\{ \frac{\lambda c_q}{2} \|f^{(\rho)} - f_z\|_{\mathcal{B}}^{q-1} \geq \epsilon \right\} \\
& \leq \text{Prob}_{z \in Z^m} \left\{ \sup_{f \in \mathcal{F}} |L_z(f)| \geq \epsilon \right\} \\
& \leq 2\mathcal{N} \left(\mathcal{F}, \frac{\epsilon}{4} \right) \exp \left(-\frac{m\epsilon^2}{8\gamma_q^2} \right).
\end{aligned} \tag{62}$$

That is,

$$\begin{aligned}
& \text{Prob}_{z \in Z^m} \left\{ \|f^{(\rho)} - f_z\|_{\mathcal{B}} \geq \epsilon \right\} \\
& \leq 2\mathcal{N} \left(\mathcal{F}, \frac{\lambda c_q \epsilon^{q-1}}{8} \right) \exp \left(-\frac{m\lambda^2 c_q^2 \epsilon^{2(q-1)}}{32\gamma_q^2} \right).
\end{aligned} \tag{63}$$

We then have (49). \square

5. Learning Rates

Proof of Theorem 3. We know from [30] that for any $f \in L^2(\rho_X)$ there holds

$$\|f - f_\rho\|_{L^2(\rho_X)}^2 = \mathcal{E}_\rho(f) - \mathcal{E}_\rho(f_\rho). \tag{64}$$

Since X is a compact set, we have by (40) that $\|f\|_{L^2(\rho_X)} \leq k\|f\|_{\mathcal{B}}$. Therefore,

$$\begin{aligned}
& \|f_z - f_\rho\|_{L^2(\rho_X)} \\
& \leq \|f_z - f^{(\rho)}\|_{L^2(\rho_X)} + \sqrt{\mathcal{E}_\rho(f^{(\rho)}) + \frac{\lambda}{q} \|f^{(\rho)}\|_{\mathcal{B}}^q - \mathcal{E}_\rho(f_\rho)} \\
& \leq k\|f_z - f^{(\rho)}\|_{\mathcal{B}} + \sqrt{\inf_{f \in \mathcal{B}} \left(\|f - f_\rho\|_{L^2(\rho_X)}^2 + \frac{\lambda}{q} \|f\|_{\mathcal{B}}^q \right)} \\
& = k\|f_z - f^{(\rho)}\|_{\mathcal{B}} + \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)}.
\end{aligned} \tag{65}$$

By (65) we have

$$\frac{1}{k} \left(\|f_z - f_\rho\|_{L^2(\rho_X)} - \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)} \right) \leq \|f_z - f^{(\rho)}\|_{\mathcal{B}}, \tag{66}$$

which gives for any $h > 0$ that

$$\begin{aligned}
& \left\{ z \in Z^m : \frac{1}{k} \left(\|f_z - f_\rho\|_{L^2(\rho_X)} - \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)} \right) > h \right\} \\
& \subset \left\{ z \in Z^m : \|f_z - f^{(\rho)}\|_{\mathcal{B}} > h \right\}.
\end{aligned} \tag{67}$$

By (49) and above inequality we have

$$\begin{aligned}
& \text{Prob}_{z \in Z^m} \left\{ \frac{1}{k} \left(\|f_z - f_\rho\|_{L^2(\rho_X)} - \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)} \right) > h \right\} \\
& \leq \text{Prob}_{z \in Z^m} \left\{ \|f_z - f^{(\rho)}\|_{\mathcal{B}} > h \right\} \\
& \leq 2\mathcal{N} \left(\mathcal{F}, \frac{\lambda c_q h^{q-1}}{8} \right) \exp \left(-\frac{m\lambda^2 c_q^2 h^{2(q-1)}}{32\gamma_q^2} \right)
\end{aligned} \tag{68}$$

or

$$\begin{aligned}
& \text{Prob}_{z \in Z^m} \left\{ \|f_z - f_\rho\|_{L^2(\rho_X)} \leq kh + \sqrt{D_q \left(f_\rho, \frac{\lambda}{q} \right)} \right\} \\
& \geq 1 - 2\mathcal{N} \left(\mathcal{F}, \frac{\lambda c_q h^{q-1}}{8} \right) \times \exp \left(-\frac{m\lambda^2 c_q^2 h^{2(q-1)}}{32\gamma_q^2} \right).
\end{aligned} \tag{69}$$

Since $\mathcal{F} \subset \mathcal{B}_1$, we know

$$\mathcal{N} \left(\mathcal{F}, \frac{\lambda c_q h^{q-1}}{8} \right) \leq \mathcal{N} \left(\mathcal{B}_1, \frac{\lambda c_q h^{q-1}}{8} \right). \tag{70}$$

By (69) and above inequality we have (16). \square

To show Theorem 4, we need two lemmas.

Lemma 12 (see [31]). *Let $c_1 > 0$, $c_2 > 0$ and $u > t > 0$. Then, the equation*

$$x^u - c_1 x^t - c_2 = 0 \tag{71}$$

has a unique positive zero x^ . In addition,*

$$x^* \leq \max \left\{ (2c_1)^{1/(u-t)}, (2c_2)^{1/u} \right\}. \tag{72}$$

Proof of Theorem 4. Since $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ has logarithmic complexity exponent $s \geq 0$, we have by (15) a constant $c_s > 0$ such that

$$\log \mathcal{N} \left(\mathcal{B}_1, \frac{\lambda c_q h^{q-1}}{8} \right) \leq c_s \left(\frac{8\gamma_q}{\lambda c_q \epsilon^{q-1}} \right)^s. \tag{73}$$

Then, by (16) we have

$$\begin{aligned} \text{Prob}_{z \in Z^m} \left\{ \|f_z - f_\rho\|_{L^2(\rho_X)} \leq k\epsilon + \sqrt{D_q(f_\rho, \lambda)} \right\} \\ \geq 1 - 2 \exp \left(c_s \left(\frac{8\gamma_q}{\lambda c_q \epsilon^{q-1}} \right)^s - \frac{m\lambda^2 c_q^2 \epsilon^{2(q-1)}}{32\gamma_q^2} \right). \end{aligned} \tag{74}$$

Take

$$2 \exp \left(c_s \left(\frac{8\gamma_q}{\lambda c_q \epsilon^{q-1}} \right)^s - \frac{m\lambda^2 c_q^2 \epsilon^{2(q-1)}}{32\gamma_q^2} \right) = \delta. \tag{75}$$

Then,

$$\begin{aligned} \epsilon^{(2+s)(q-1)} - \frac{32\gamma_q^2 \log(2/\delta)}{m\lambda^2 c_q^2} \epsilon^{s(q-1)} \\ - \frac{32\gamma_q^2 c_s}{m\lambda^2 c_q^2} \times \left(\frac{8\gamma_q}{\lambda c_q} \right)^s = 0. \end{aligned} \tag{76}$$

By Lemma 12, we know that the unique solution ϵ^* of (75) satisfies

$$\begin{aligned} \epsilon^* \leq \max \left\{ \left(\frac{64\gamma_q^2 \log(2/\delta)}{m\lambda^2 c_q^2} \right)^{1/2(q-1)}, \right. \\ \left. \left(\frac{64\gamma_q^2 c_s}{m\lambda^2 c_q^2} \times \left(\frac{8\gamma_q}{\lambda c_q} \right)^s \right)^{1/(2+s)(q-1)} \right\} \\ \leq \left(\frac{64\gamma_q^2 \log(2/\delta)}{m\lambda^2 c_q^2} \right)^{1/2(q-1)} \\ + \left(\frac{8^{2+s} \gamma_q^{2+s} c_s}{m\lambda^{2+s} c_q^{2+s}} \right)^{1/(2+s)(q-1)}. \end{aligned} \tag{77}$$

By (74) and (77), we have (19). □

Acknowledgments

This work was supported partially by the National Natural Science Foundation of China under Grant nos. 10871226, 61179041, 11271199. The authors thank the reviewers for giving many valuable suggestions and comments which make the paper presented in a better form.

References

- [1] S. Loustau, "Aggregation of SVM classifiers using Sobolev spaces," *Journal of Machine Learning Research*, vol. 9, pp. 1559–1582, 2008.
- [2] C. A. Micchelli and M. Pontil, "A function representation for learning in Banach spaces," in *Learning Theory*, vol. 3120 of *Lecture Notes on Computer Science*, pp. 255–269, Springer, Berlin, Germany, 2004.
- [3] S. G. Lv and J. D. Zhu, "Error bounds for l^p -norm multiple kernel learning with least square loss," *Abstract and Applied Analysis*, vol. 2012, Article ID 915920, 18 pages, 2012.
- [4] H. Zhang, Y. Xu, and J. Zhang, "Reproducing kernel Banach spaces for machine learning," *Journal of Machine Learning Research*, vol. 10, pp. 2741–2775, 2009.
- [5] H. Zhang and J. Zhang, "Regularized learning in Banach spaces as an optimization problem: representer theorems," *Journal of Global Optimization*, vol. 54, no. 2, pp. 235–250, 2012.
- [6] H. Zhang and J. Zhang, "Generalized semi-inner products with applications to regularized learning," *Journal of Mathematical Analysis and Applications*, vol. 372, no. 1, pp. 181–196, 2010.
- [7] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [8] M. Z. Nashed and Q. Sun, "Sampling and reconstruction of signals in a reproducing kernel subspace of $L_p(\mathfrak{R}^d)$," *Journal of Functional Analysis*, vol. 258, no. 7, pp. 2422–2452, 2010.
- [9] F. H. Clarke, Y. S. Ledyae, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis And Control Theory*, vol. 178 of *Graduate Texts in Mathematics*, Springer, Berlin, Germany, 1998.
- [10] H. K. Xu, "Inequalities in Banach spaces with applications," *Nonlinear Analysis. Theory, Methods & Applications*, vol. 16, no. 12, pp. 1127–1138, 1991.
- [11] Z. B. Xu and G. F. Roach, "Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces," *Journal of Mathematical Analysis and Applications*, vol. 157, no. 1, pp. 189–210, 1991.
- [12] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, "Minimization of Tikhonov functionals in Banach spaces," *Abstract and Applied Analysis*, vol. 2008, Article ID 192679, 18 pages, 2008.
- [13] Z. B. Xu and Z. S. Zhang, "Another set of characteristic inequalities of L^p Banach spaces," *Acta Mathematica Sinica*, vol. 37, no. 4, pp. 433–439, 1994 (Chinese).
- [14] K. S. Kazimierski, "Minimization of the Tikhonov functional in Banach spaces smooth and convex of power type by steepest descent in the dual," *Computational Optimization and Applications*, vol. 48, no. 2, pp. 309–324, 2011.
- [15] F. Cucker and D. X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, vol. 24 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, New York, NY, USA, 2007.

- [16] B. H. Sheng, J. L. Wang, and P. Li, "The covering number for some Mercer kernel Hilbert spaces," *Journal of Complexity*, vol. 24, no. 2, pp. 241–258, 2008.
- [17] B. H. Sheng, J. L. Wang, and Z. X. Chen, "The covering number for some Mercer kernel Hilbert spaces on the unit sphere," *Taiwanese Journal of Mathematics*, vol. 15, no. 3, pp. 1325–1340, 2011.
- [18] H. W. Sun and D. X. Zhou, "Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels," *Journal of Fourier Analysis and Applications*, vol. 14, no. 1, pp. 89–101, 2008.
- [19] D. X. Zhou, "The covering number in learning theory," *Journal of Complexity*, vol. 18, no. 3, pp. 739–767, 2002.
- [20] C. Gansner, "Modulus of continuity conditions for Jacobi series," *Journal of Mathematical Analysis and Applications*, vol. 27, no. 3, pp. 575–600, 1969.
- [21] H. Berens and Y. Xu, "On Bernstein-Durrmeyer polynomials with Jacobi-weights," in *Approximation Theory and Functional Analysis*, C. K. Chui, Ed., pp. 25–46, Academic Press, Boston, Mass, USA, 1991.
- [22] E. E. Berdysheva and K. Jetter, "Multivariate Bernstein-Durrmeyer operators with arbitrary weight functions," *Journal of Approximation Theory*, vol. 162, no. 3, pp. 576–598, 2010.
- [23] E. E. Berdysheva, "Uniform convergence of Bernstein-Durrmeyer operators with respect to arbitrary measure," *Journal of Mathematical Analysis and Applications*, vol. 394, no. 1, pp. 324–336, 2012.
- [24] F. Filbir and H. N. Mhaskar, "Marcinkiewicz-Zygmund measures on manifolds," *Journal of Complexity*, vol. 27, no. 6, pp. 568–596, 2011.
- [25] D. X. Zhou and K. Jetter, "Approximation with polynomial kernels and SVM classifiers," *Advances in Computational Mathematics*, vol. 25, no. 1–3, pp. 323–344, 2006.
- [26] H. Z. Tong, D. R. Chen, and L. Z. Peng, "Learning rates for regularized classifiers using multivariate polynomial kernels," *Journal of Complexity*, vol. 24, no. 5–6, pp. 619–631, 2008.
- [27] B. Z. Li, "Approximation by multivariate Bernstein-Durrmeyer operators and learning rates of least-squares regularized regression with multivariate polynomial kernels," *Journal of Approximation Theory*, vol. 173, pp. 33–55, 2013.
- [28] A. Guntuboyina and B. Sen, "Covering numbers for convex functions," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 1957–1965, 2013.
- [29] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research and Financial Engineering, Springer, New York, NY, USA, 2000.
- [30] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society*, vol. 39, no. 1, pp. 1–49, 2002.
- [31] F. Cucker and S. Smale, "Best choices for regularization parameters in learning theory: on the bias-variance problem," *Foundations of Computational Mathematics*, vol. 2, no. 4, pp. 413–428, 2002.