

Research Article

Coefficient-Based Regression with Non-Identical Unbounded Sampling

Jia Cai

School of Mathematics and Computational Science, Guangdong University of Business Studies, Guangzhou, Guangdong 510320, China

Correspondence should be addressed to Jia Cai; jiacai1999@gmail.com

Received 18 January 2013; Accepted 15 April 2013

Academic Editor: Qiang Wu

Copyright © 2013 Jia Cai. 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 investigate a coefficient-based least squares regression problem with indefinite kernels from non-identical unbounded sampling processes. Here non-identical unbounded sampling means the samples are drawn independently but not identically from unbounded sampling processes. The kernel is not necessarily symmetric or positive semi-definite. This leads to additional difficulty in the error analysis. By introducing a suitable reproducing kernel Hilbert space (RKHS) and a suitable intermediate integral operator, elaborate analysis is presented by means of a novel technique for the sample error. This leads to satisfactory results.

1. Introduction and Preliminary

We study coefficient-based least squares regression with indefinite kernels from non-identical unbounded sampling processes. In our setting, functions are defined on a compact subset X of \mathbb{R}^n and take values in $Y = \mathbb{R}$. Let ρ be a Borel probability measure on $Z = X \times Y$. A sample $\mathbf{z} = \{z_t = (x_t, y_t)\}_{t=1}^T \subset Z^T$ is drawn independently from different Borel probability measures $\rho^{(t)}$ ($t = 1, \dots, T$), $\rho^{(t)}(\cdot | x) = \rho(\cdot | x)$. Let $\rho_X^{(t)}$ be the marginal distribution of $\rho^{(t)}$ on X and ρ_X the marginal distribution of ρ on X . We assume that the sequence $\{\rho_X^{(t)}\}$ converges exponentially fast in the dual of the Hölder space $C^s(X)$. Here the Hölder space $C^s(X)$ ($0 \leq s \leq 1$) is defined as the space of all continuous functions on X with the following norm finite [1]:

$$\|f\|_{C^s(X)} = \|f\|_{\infty} + |f|_{C^s(X)}, \quad (1)$$

where

$$|f|_{C^s(X)} := \sup_{x \neq y \in X} \frac{|f(x) - f(y)|}{(d(x, y))^s}. \quad (2)$$

Definition 1. Let $0 \leq s \leq 1$; we say that the sequence $\{\rho_X^{(t)}\}$ converges exponentially fast in $(C^s(X))^*$ to a probability

measure ρ_X on X or converges exponentially in short if there exist $C_1 > 0$ and $0 < \alpha < 1$ such that

$$\|\rho_X^{(t)} - \rho_X\|_{(C^s(X))^*} \leq C_1 \alpha^t, \quad \forall t \in \mathbb{N}. \quad (3)$$

By the definition of the dual space $(C^s(X))^*$, the decay condition (3) can be expressed as

$$\left| \int_X f(x) d\rho_X^{(t)} - \int_X f(x) d\rho_X \right| \leq C_1 \alpha^t \|f\|_{C^s(X)}, \quad (4)$$

$$\forall f \in C^s(X), t \in \mathbb{N}.$$

The regression function $f_\rho : X \rightarrow Y$ is given by

$$f_\rho(x) = \int_Y y d\rho(y | x), \quad x \in X, \quad (5)$$

where $\rho(y | x)$ is the conditional distribution of y at $x \in X$ and ρ is unknown, so f_ρ cannot be obtained directly. The aim of regression problem is to learn a good approximation of f_ρ from sample \mathbf{z} . This is an ill-posed problem and regularization scheme is needed.

Classical learning algorithm is conducted by a scheme in a reproducing kernel Hilbert space (RKHS) [2] associated with a Mercer kernel $K : X \times X \rightarrow \mathbb{R}$, which is defined to be a continuous, symmetric, and positive semi-definite (p.s.d.)

function. RKHS \mathcal{H}_K is defined to be the completion of the linear span of $\{K_x = K(\cdot, x) : x \in X\}$ with the inner product $\langle K_x, K_y \rangle_K = K(x, y)$. Define $\kappa = \sup_{x, y \in X} |K(x, y)| < \infty$; then the regularized regression problem is given by

$$f_{\mathbf{z}} = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{T} \sum_{t=1}^T (f(x_t) - y_t)^2 + \lambda \|f\|_K^2 \right\}. \quad (6)$$

It has been well understood due to lots of the literature ([3–5] and the references therein). Here we consider the indefinite kernel scheme in a hypothesis space $\mathcal{H}_{K, \mathbf{z}}$ depending on the sample \mathbf{z} ; this space is defined by

$$\mathcal{H}_{K, \mathbf{z}} = \left\{ \sum_{i=1}^T \beta_i K_{x_i} : \beta_i \in \mathbb{R} \right\}. \quad (7)$$

And the regularized penalty term is imposed on the coefficients of function f . Indefinite kernel K means K does not need to satisfy symmetry and p.s.d. condition except for continuity and boundedness. Define $\bar{K}(u, v) = \int_X K(u, x)K(v, x)d\rho_X(x)$; then $\bar{K}(u, v)$ is a Mercer kernel. For more introductions about learning with indefinite kernels, please see [6–8]. For all $x \in X$, if we define $L_{K, \rho_X} f(x) = \int_X K(x, u)f(u)d\rho_X(u)$ and $L_{K, \rho_X^{(t)}} f(x) = \int_X K(x, u)f(u)d\rho_X^{(t)}(u)$, since X is compact and K is continuous, L_{K, ρ_X} and its adjoint L_{K, ρ_X}^* are both compact operators. Hence $L_{\bar{K}, \rho_X} = L_{K, \rho_X} L_{K, \rho_X}^*$, $L_{\bar{K}, \rho_X^{(t)}} = L_{K, \rho_X^{(t)}} L_{K, \rho_X^{(t)}}^*$. The learning algorithm we are interested in this paper takes the following form:

$$f_{\mathbf{z}, \lambda} = \arg \min_{f \in \mathcal{H}_{K, \mathbf{z}}} \left\{ \frac{1}{T} \sum_{i=1}^T (f(x_i) - y_i)^2 + \lambda \Omega_{\mathbf{z}}(f) \right\}, \quad \lambda > 0. \quad (8)$$

We define the following coefficient-based regularizer:

$$\Omega_{\mathbf{z}}(f) = \|f\|_{\mathbf{z}}^2 := T \sum_{i=1}^T \beta_i^2 \quad \text{for } f = \sum_{i=1}^T \beta_i K_{x_i}. \quad (9)$$

Then we have $f_{\mathbf{z}, \lambda} = f_{\beta_{\mathbf{z}, \lambda}} = \sum_{i=1}^T \beta_{\mathbf{z}, \lambda, i} K_{x_i}$. By using the integral operator technique from [4], in [9], Sun and Wu gave the capacity independent estimate for the convergence rate of $\|f_{\mathbf{z}, \lambda} - f_{\rho}\|_{\rho}^2$, where $\|f\|_{\rho} = (\int_X |f(x)|^2 d\rho_X(x))^{1/2}$. Shi investigated the error analysis in a data dependent hypothesis space for general kernels [10]. Sun and Guo conducted error analysis for the Mercer kernels by uniform bounded non-i.i.d. sampling [11]. In this paper, we study learning algorithm (8) by non-identical unbounded sampling processes with indefinite kernels.

If K is a Mercer kernel, from [3] we know that \mathcal{H}_K is in the range of $L_K^{1/2}$. For an indefinite kernel K , recall $L_{\bar{K}, \rho_X} = L_{K, \rho_X} L_{K, \rho_X}^*$. Based on the polar decomposition of compact operators ([12]).

Lemma 2. *Let H be a separable Hilbert space and T a compact operator on H ; then T can be factored as*

$$T = \Gamma A, \quad (10)$$

where $A = (T^* T)^{1/2}$ and Γ is a partial isometry on H with $\Gamma^* \Gamma$ being orthogonal projection onto $\overline{\mathcal{R}(A)}$.

We immediately have the following proposition [10].

Proposition 3. *Consider $\mathcal{H}_{\bar{K}}$ as a subspace of $L_{\rho_X}^2$; then $L_K^* = UL_{\bar{K}}^{1/2}$ and $L_K = L_{\bar{K}}^{1/2} U^*$, where U is a partial isometry on $L_{\rho_X}^2$ with $U^* U$ being the orthogonal projection onto $\overline{\mathcal{H}_{\bar{K}}}$.*

We use the RKHS $\mathcal{H}_{\bar{K}}$ to approximate f_{ρ} , hence define

$$f_{\lambda, \rho_X} = \arg \min_{f \in \mathcal{H}_{\bar{K}}} \left\{ \int_Z (f(x) - f_{\rho}(x))^2 d\rho(x, y) + \lambda \|f\|_{\bar{K}}^2 \right\}. \quad (11)$$

In order to estimate $f_{\mathbf{z}, \lambda} - f_{\rho}$, we construct

$$f_{\lambda, \bar{\rho}_X^{(T)}} = \left(\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}} \right)^{-1} L_{\bar{K}, \bar{\rho}_X^{(T)}} f_{\rho}, \quad (12)$$

where $\bar{\rho}_X^{(T)} = (1/T) \sum_{t=1}^T \rho_X^{(t)}$. Then we can decompose the error term into the following three parts:

$$\begin{aligned} f_{\mathbf{z}, \lambda} - f_{\rho} &= f_{\mathbf{z}, \lambda} - f_{\lambda, \bar{\rho}_X^{(T)}} \\ &\quad + f_{\lambda, \bar{\rho}_X^{(T)}} - f_{\lambda, \rho_X} + f_{\lambda, \rho_X} - f_{\rho}. \end{aligned} \quad (13)$$

We will conduct the error analysis in several steps. The major contribution we make is on the sample error estimate; the main difficulty is the non-identical unbounded sampling of the samples; we overcome this difficulty by introducing a suitable intermediate operator.

2. Key Analysis and Main Results

In order to give the error analysis, we assume that the kernel \bar{K} satisfies the following kernel condition [1, 11].

Definition 4. We say that the Mercer kernel \bar{K} satisfies the kernel condition of order s if, for some constant $\kappa_s > 0$, $\bar{K} \in C^s(X \times X)$, and for all $u, v \in X$,

$$\|\bar{K}_u - \bar{K}_v\|_{\bar{K}} \leq \kappa_s (d(u, v))^s. \quad (14)$$

Since sample \mathbf{z} is drawn from unbounded sampling processes, we will assume the following moment hypothesis condition [13].

Moment Hypothesis. There exist constants $M > 0$ and $C_2 > 0$ such that

$$\int_Y |y|^l d\rho(y | x) \leq C_2 l! M^l \quad \forall l \in \mathbb{N}, x \in X. \quad (15)$$

There is a large literature on error analysis for learning algorithm (6); see, for example, [4, 5, 14–17]. But most obtained results are presented under the standard assumption that $|y| \leq M$ almost surely (M is a constant). This excludes the case of Gaussian noise. The moment hypothesis condition

is a natural generalization of the condition $|y| \leq M$. Wang and Zhou considered error analysis for algorithm (6) under condition (15). Our main results are about learning rates of algorithm (8) under conditions (3), (14) and the approximation ability of $\mathcal{H}_{\bar{K}}$ in terms of f_ρ .

Now we can state our general results on learning rates for algorithm (8).

Theorem 5. Assume moment hypothesis condition (15); $\rho_X^{(t)}$ satisfies condition (3) and \bar{K} satisfies condition (14); $f_\rho \in L_{\bar{K}, \rho_X}^r(L_{\rho_X}^2)$, for $1/2 < r \leq 3/2$; take $\lambda = T^{-\theta}$ with $0 < \theta < 1/3$; then

$$\mathbb{E} \|f_{z,\lambda} - f_\rho\|_\rho \leq C_\kappa'' T^{-\min\{1/2-(3/2)\theta, (r-1/2)\theta\}}, \quad (16)$$

where C_κ'' is a constant depending on κ, s , and α , but not on T or δ , and will be given explicitly in Section 3.3.

Remark 6. If we take $\lambda = T^{-1/2(r+1)}$, then our rate is $O(T^{-(2r-1)/4(r+1)})$. The proof of Theorem 5 will be conducted in Section 3, where the error term is decomposed into three parts. In [11], the authors consider the coefficient-based regression with the Mercer kernels by uniform bounded non-i.i.d sampling; the best rate of order $O(T^{-2r/(1+2r)})$ was obtained.

When the samples are drawn i.i.d from measure ρ , we have the following result.

Theorem 7. Assume moment hypothesis condition (15); \bar{K} satisfies condition (14); $f_\rho \in L_{\bar{K}, \rho_X}^r(L_{\rho_X}^2)$; then if $0 < r \leq 1$, take $\lambda = T^{-1/(2r+3)}$; one see that

$$\mathbb{E} \|f_{z,\lambda} - f_\rho\|_\rho \leq \bar{C}_\kappa T^{-r/(2r+3)}, \quad (17)$$

where \bar{C}_κ is a constant depending on κ, s , and α but not on T or δ . And if $r > 1$, take $\lambda = T^{-1/5}$; we have

$$\mathbb{E} \|f_{z,\lambda} - f_\rho\|_\rho \leq \bar{C}_\kappa T^{-1/5}. \quad (18)$$

Here we get the same learning rate as the one in [9]. But our rate is derived under a relaxation condition of the sampling output.

3. Error Analysis

In this section, we will state the error analysis in several steps.

3.1. Regularization Error Estimation. In this subsection, we address a bound for the regularization error $\|f_{\lambda, \rho_X} - f_\rho\|$. The error estimate for regularization error has been investigated in lots of the literature in learning theory ([4, 18] and the references therein); we will omit the proof and quote it directly.

Proposition 8. Assume $f_\rho = L_{\bar{K}, \rho_X}^r(g_\rho)$ for some $g_\rho \in L_{\rho_X}^2$ and $r > 0$; the following bound for approximation error holds:

$$\|f_{\lambda, \rho_X} - f_\rho\|_\rho \leq C_q \lambda^q, \quad (19)$$

where $C_q = (1 + \kappa^{2r-2}) \|L_{\bar{K}, \rho_X}^{-r} f_\rho\|_\rho$, $q = \min\{r, 1\}$, and when $1/2 < r \leq 3/2$,

$$\|f_{\lambda, \rho_X} - f_\rho\|_{\bar{K}} \leq C_r \lambda^{r-1/2}, \quad (20)$$

where $C_r = \|L_{\bar{K}, \rho_X}^{-r} f_\rho\|_\rho$.

3.2. Estimate for the Measure Error. This subsection is devoted to the analysis of the term $\|f_{\lambda, \rho_X^{(T)}} - f_{\lambda, \rho_X}\|_{\bar{K}}$ caused by the difference of measures, which we called measure error. The ideas of proof are from [1]. Before giving the result, let us state a crucial lemma first.

Lemma 9. Assume \bar{K} satisfies condition (14); then

$$\|L_{\bar{K}, \rho_X^{(T)}} - L_{\bar{K}, \rho_X}\| \leq (\kappa + \kappa_s)^2 \|\rho_X - \bar{\rho}_X^{(T)}\|_{(C^s(X))^*}. \quad (21)$$

Proof. For any $h \in C^s(X)$, we see that

$$\begin{aligned} & \left\| \left(L_{\bar{K}, \rho_X^{(T)}} - L_{\bar{K}, \rho_X} \right) h \right\|_{\bar{K}}^2 \\ &= \int_X h(u) \left\{ \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right\} \\ & \quad \times d(\rho_X - \bar{\rho}_X^{(T)})(u) \\ & \leq \|\rho_X - \bar{\rho}_X^{(T)}\|_{(C^s(X))^*} \|g\|_{C^s(X)} \\ & = \|\rho_X - \bar{\rho}_X^{(T)}\|_{(C^s(X))^*} (\|g\|_\infty + |g|_{C^s(X)}), \end{aligned} \quad (22)$$

where $g(u) = h(u) \left\{ \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right\}$, $u \in X$. Now we need to estimate $\|g\|_\infty$ and $|g|_{C^s(X)}$, respectively. For the term $\|g\|_\infty$, it is easy to see that

$$\|g\|_\infty \leq \|h\|_\infty \sup_{u \in X} \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right|. \quad (23)$$

The estimation of $|g|_{C^s(X)}$ is more involved:

$$\begin{aligned} |g|_{C^s(X)} & \leq |h|_{C^s(X)} \sup_{u \in X} \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right| \\ & \quad + \|h\|_\infty \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right|_{C^s(X)}. \\ & \sup_{u \in X} \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right| \\ & \leq \|\rho_X - \bar{\rho}_X^{(T)}\|_{(C^s(X))^*} \|h(\cdot) \bar{K}(u, \cdot)\|_{C^s(X)} \\ & \leq \|\rho_X - \bar{\rho}_X^{(T)}\|_{(C^s(X))^*} \\ & \quad \times \{ \kappa^2 \|h\|_\infty + |h|_{C^s(X)} \kappa^2 + \|h\|_\infty \kappa \kappa_s \}, \end{aligned}$$

$$\begin{aligned}
& \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right|_{C^s(X)} \\
&= \sup_{u_1 \neq u_2 \in X} \left| \int_X h(v) \frac{\bar{K}(u_1, v) - \bar{K}(u_2, v)}{d(u_1, u_2)^s} \right. \\
&\quad \left. \times d(\rho_X - \bar{\rho}_X^{(T)})(v) \right| \\
&\leq \sup_{u_1 \neq u_2 \in X} \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*} \\
&\quad \times \left\| h(v) \frac{\bar{K}(u_1, v) - \bar{K}(u_2, v)}{d(u_1, u_2)^s} \right\|_{C^s(X)}. \tag{24}
\end{aligned}$$

Since

$$\begin{aligned}
& \sup_{u_1 \neq u_2 \in X} \left\| h(v) \frac{\bar{K}(u_1, v) - \bar{K}(u_2, v)}{d(u_1, u_2)^s} \right\|_{C^s(X)} \\
&\leq \|h\|_\infty \sup_{u_1 \neq u_2 \in X} \sup_{v \in X} \left| \frac{\bar{K}(u_1, v) - \bar{K}(u_2, v)}{d(u_1, u_2)^s} \right| \\
&\quad + |h|_{C^s(X)} \sup_{u_1 \neq u_2 \in X} \sup_{v \in X} \left| \frac{\bar{K}(u_1, v) - \bar{K}(u_2, v)}{d(u_1, u_2)^s} \right| \\
&\quad + \|h\|_\infty \sup_{u_1 \neq u_2 \in X} \sup_{v_1 \neq v_2 \in X} \left(|\bar{K}(u_1, v_2) - \bar{K}(u_2, v_2) \right. \\
&\quad \quad \left. - \bar{K}(u_1, v_1) + \bar{K}(u_2, v_1) \right| \\
&\quad \quad \times (d(u_1, u_2)^s d(v_1, v_2)^s)^{-1} \Big) \\
&\leq \|h\|_{C^s(X)} \kappa \kappa_s + \|h\|_\infty \kappa_s^2, \tag{25}
\end{aligned}$$

then

$$\begin{aligned}
& \left| \int_X h(v) \bar{K}(u, v) d(\rho_X - \bar{\rho}_X^{(T)})(v) \right|_{C^s(X)} \\
&\leq \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*} \|h\|_{C^s(X)} (\kappa \kappa_s + \kappa_s^2). \tag{26}
\end{aligned}$$

Therefore

$$\begin{aligned}
|g|_{C^s(X)} &\leq \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*} \\
&\quad \times \left[|h|_{C^s(X)} \{ \kappa^2 \|h\|_{C^s(X)} + \|h\|_\infty \kappa \kappa_s \} \right. \\
&\quad \left. + \|h\|_\infty \|h\|_{C^s(X)} (\kappa \kappa_s + \kappa_s^2) \right]. \tag{27}
\end{aligned}$$

Combining the estimation of $\|g\|_\infty$ and $|g|_{C^s(X)}$, we get

$$\begin{aligned}
& \left\| (L_{\bar{\kappa}, \bar{\rho}_X^{(T)}} - L_{\bar{\kappa}, \rho_X}) h \right\|_{\bar{\kappa}}^2 \\
&\leq \|h\|_{C^s(X)}^2 \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*}^2 (\kappa + \kappa_s)^2. \tag{28}
\end{aligned}$$

When condition (14) is satisfied, it was proved in [19] that $\mathcal{H}_{\bar{\kappa}}$ is included in $C^s(X)$ with the inclusion bounded

$$\|f\|_{C^s(X)} \leq (\kappa + \kappa_s) \|f\|_{\bar{\kappa}}, \quad \forall f \in \mathcal{H}_{\bar{\kappa}}. \tag{29}$$

Then

$$\left\| (L_{\bar{\kappa}, \bar{\rho}_X^{(T)}} - L_{\bar{\kappa}, \rho_X}) h \right\|_{\bar{\kappa}} \leq \|h\|_{\bar{\kappa}} (\kappa + \kappa_s)^2 \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*}. \tag{30}$$

This completes the proof. \square

Proposition 10. Assume $f_\rho \in L_{\bar{\kappa}, \rho_X}^r(L_{\rho_X}^2)$ for some $1/2 < r \leq 3/2$; \bar{K} satisfies condition (14); the following bound for measure error holds:

$$\left\| f_{\lambda, \bar{\rho}_X^{(T)}} - f_{\lambda, \rho_X} \right\|_{\bar{\kappa}} \leq \frac{C_3 \lambda^{r-3/2}}{T}, \tag{31}$$

where $C_\kappa = (\kappa + \kappa_s^2)^2$ and $C_3 = C_1 C_\kappa C_r \alpha / (1 - \alpha)$.

Proof. From (11), simple calculation shows that $f_{\lambda, \rho_X} = (\lambda I + L_{\bar{\kappa}, \rho_X})^{-1} L_{\bar{\kappa}, \rho_X} f_\rho$. Recalling (12), we can see that

$$\begin{aligned}
& \left\| f_{\lambda, \bar{\rho}_X^{(T)}} - f_{\lambda, \rho_X} \right\|_{\bar{\kappa}} \\
&= \left\| (\lambda I + L_{\bar{\kappa}, \bar{\rho}_X^{(T)}})^{-1} \left\{ (L_{\bar{\kappa}, \bar{\rho}_X^{(T)}} - L_{\bar{\kappa}, \rho_X}) f_\rho \right. \right. \\
&\quad \left. \left. + (L_{\bar{\kappa}, \rho_X} - L_{\bar{\kappa}, \bar{\rho}_X^{(T)}}) f_{\lambda, \rho_X} \right\} \right\|_{\bar{\kappa}} \\
&= \left\| (\lambda I + L_{\bar{\kappa}, \bar{\rho}_X^{(T)}})^{-1} (L_{\bar{\kappa}, \bar{\rho}_X^{(T)}} - L_{\bar{\kappa}, \rho_X}) (f_\rho - f_{\lambda, \rho_X}) \right\|_{\bar{\kappa}} \\
&\leq \frac{1}{\lambda} \left\| (L_{\bar{\kappa}, \bar{\rho}_X^{(T)}} - L_{\bar{\kappa}, \rho_X}) (f_\rho - f_{\lambda, \rho_X}) \right\|_{\bar{\kappa}}. \tag{32}
\end{aligned}$$

Applying Lemma 9 to the case $h = f_\rho - f_{\lambda, \rho_X}$, we get

$$\begin{aligned}
& \left\| f_{\lambda, \bar{\rho}_X^{(T)}} - f_{\lambda, \rho_X} \right\|_{\bar{\kappa}} \\
&\leq \frac{1}{\lambda} (\kappa + \kappa_s)^2 \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*} \|f_{\lambda, \rho_X} - f_\rho\|_{\bar{\kappa}}. \tag{33}
\end{aligned}$$

By the definition of $\bar{\rho}_X^{(T)}$ and noticing (3), we can see

$$\begin{aligned}
& \left\| \rho_X - \bar{\rho}_X^{(T)} \right\|_{(C^s(X))^*} \\
&= \frac{1}{T} \sum_{t=1}^T \left\| \rho_X^{(t)} - \rho_X \right\|_{(C^s(X))^*} \leq \frac{C_1 \alpha}{T(1 - \alpha)}. \tag{34}
\end{aligned}$$

This in connection with Proposition 8 yields the conclusion. \square

3.3. Sample Error Estimation. In this subsection we will conduct the estimation of the term $f_{z, \lambda} - f_{\lambda, \bar{\rho}_X^{(T)}}$. At first, we give some notations. Let $C(X)$ be the space of bounded continuous functions on X with supremum norm $\|\cdot\|_\infty$.

Define sampling operator $S_x : C(X) \rightarrow \mathbb{R}^T$ by $S_x(f) = (f(x_1), \dots, f(x_T))$ ([18]). For $\beta = (\beta_1, \dots, \beta_T)$, let U_x and \widehat{U}_x be operators from \mathbb{R}^T to $C(X)$ defined as

$$U_x \beta = \frac{1}{T} \sum_{i=1}^T \beta_i K(x, x_i), \quad \widehat{U}_x \beta = \frac{1}{T} \sum_{i=1}^T \beta_i K(x_i, x). \quad (35)$$

It is easy to see that both U_x and \widehat{U}_x are bounded operators. Recall the definition of $f_{z,\lambda}$; then

$$\beta_{z,\lambda} = \arg \min_{\beta \in \mathbb{R}^T} \left\{ \frac{1}{T} \sum_{i=1}^T (f_\beta(x_i) - y_i)^2 + \lambda T \sum_{i=1}^T \beta_i^2 \right\}. \quad (36)$$

Computing the gradient of the above equation, we immediately have [9]

$$\beta_{z,\lambda} = \frac{1}{T} (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y. \quad (37)$$

Hence $f_{z,\lambda} = U_x (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y$. Employing the method as shown in [9], we can decompose the sample error into two parts:

$$\begin{aligned} f_{z,\lambda} - f_{\lambda, \bar{\rho}_X^{(T)}} &= U_x (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y \\ &\quad - (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho \\ &= \left\{ U_x (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y \right. \\ &\quad \left. - (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} U_x S_x \widehat{U}_x y \right\} \\ &\quad + \left\{ (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} U_x S_x \widehat{U}_x y \right. \\ &\quad \left. - (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho \right\} \\ &= \text{I} + \text{II}. \end{aligned} \quad (38)$$

Now we state our estimation for the sample error. The estimates are more involved since the sample is drawn by non-identical unbounded sampling processes. We overcome the difficulty by introducing a stepping integral operator $L_{\bar{K}, \bar{\rho}_X^{(T)}}$ which plays an intermediate role in the estimates, and the definition of it will be given later.

Theorem 11. *Let $f_{z,\lambda}$ be given by (8), assume moment hypothesis condition (15), and the marginal distribution sequence $\rho_X^{(t)}$, $t \in \mathbb{N}$, satisfies condition (3); then*

$$\mathbb{E} \|f_{z,\lambda} - f_{\lambda, \bar{\rho}_X^{(T)}}\|_\rho \leq \frac{C'_\kappa}{\lambda^{3/2} T^{1/2}}, \quad (39)$$

where $C_4 = \alpha C_1 \kappa M (\kappa + 2|K|_{C^1(X \times X)}) (\kappa \sqrt{2C_2} + C_2)$, $C'_\kappa = 4\sqrt{3C_2} M \kappa^2 + 4\sqrt{5C_2} M \kappa^3 + C_4 / (1 - \alpha)$.

Proof. We will estimate I and II, respectively:

$$\begin{aligned} \text{I} &= (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}}) \\ &\quad \times U_x (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y \\ &\quad - (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} U_x (\lambda I + S_x \widehat{U}_x S_x U_x) \\ &\quad \times (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y \\ &= (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} (L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x) \\ &\quad \times (\lambda I + S_x \widehat{U}_x S_x U_x)^{-1} S_x \widehat{U}_x y \\ &= (\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}})^{-1} (L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x) T \beta_{z,\lambda}. \end{aligned} \quad (40)$$

Then

$$\begin{aligned} \|\text{I}\|_\rho &\leq \lambda^{-1} T \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\| \|\beta_{z,\lambda}\|_2 \\ &\quad + \lambda^{-1} T \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x\| \|\beta_{z,\lambda}\|_2, \end{aligned} \quad (41)$$

where $\widehat{K}(x, u) = \int_X K(x, v) K(u, v) d\bar{\rho}_X^{(T)}(v)$ and $\|\beta_{z,\lambda}\|_2$ is the l^2 norm on \mathbb{R}^T , for $\|\beta_{z,\lambda}\|_2$; noticing (36), we can have

$$\lambda T \|\beta_{z,\lambda}\|_2^2 \leq \frac{1}{T} \sum_{i=1}^T y_i^2. \quad (42)$$

This means

$$\begin{aligned} \mathbb{E} \|\text{I}\|_\rho &\leq \lambda^{-1} T (\mathbb{E} \|\beta_{z,\lambda}\|_2^2)^{1/2} \left\{ \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\|^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x\|^2 \right)^{1/2} \right\} \\ &\leq \lambda^{-3/2} \sqrt{T} \left(\mathbb{E} \left(\frac{1}{T} \sum_{i=1}^T y_i^2 \right) \right)^{1/2} \\ &\quad \times \left\{ \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\|^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x\|^2 \right)^{1/2} \right\} \\ &\leq \lambda^{-3/2} \sqrt{2C_2 T M} \left\{ \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\|^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x\|^2 \right)^{1/2} \right\}. \end{aligned} \quad (43)$$

According to the definition of U_x , for any $\beta \in \mathbb{R}^T$, $\|U_x \beta\|_\rho \leq (\kappa/\sqrt{T})\|\beta\|_2$; this implies that $\|U_x\| \leq \kappa/\sqrt{T}$. Therefore

$$\begin{aligned} & \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} U_x\| \\ &= \sup_{\|\beta\|_2 \leq 1} \left\| \frac{1}{T} \sum_{i=1}^T \beta_i \int_X (\bar{K}(x, u) - \widehat{K}(x, u)) \right. \\ & \quad \times K(u, x_i) d\bar{\rho}_X^{(T)}(u) \left. \right\|_\rho \\ & \leq \sup_{\|\beta\|_2 \leq 1} \frac{1}{T} \sum_{i=1}^T |\beta_i| \left\| \left(\int_X K^2(u, x_i) d\bar{\rho}_X^{(T)}(u) \right)^{1/2} \right. \\ & \quad \times \left(\int_X [\bar{K}(x, u) - \widehat{K}(x, u)]^2 \right. \\ & \quad \left. \left. \times d\bar{\rho}_X^{(T)}(u) \right)^{1/2} \right\|_\rho \\ & \leq \frac{\kappa}{\sqrt{T}} \left\| \int_X K(x, v) K(u, v) d(\bar{\rho}_X^{(T)} - \rho_X)(v) \right\|_\infty \\ & \leq \frac{\kappa}{\sqrt{T}} \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^*(X))^*} \|K(x, \cdot) K(u, \cdot)\|_{C^*(X)} \\ & \leq \frac{\alpha C_1 \kappa^2 (\kappa + 2|K|_{C^*(X \times X)})}{T^{3/2} (1 - \alpha)}. \end{aligned} \tag{44}$$

Hence

$$\begin{aligned} & \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} U_x\|^2 \right)^{1/2} \\ & \leq \frac{\alpha C_1 \kappa^2 (\kappa + 2|K|_{C^*(X \times X)})}{T^{3/2} (1 - \alpha)}. \end{aligned} \tag{45}$$

For the term $\|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\|$, let

$$\begin{aligned} \eta_{t,j,l}(x) &= K(x_t, x_l) K(x_t, x_j) K(x, x_j) \\ & \quad - \iint_X K(x, v) K(u, v) K(u, x_l) d\rho_X^{(j)}(v) d\rho_X^{(t)}(u), \end{aligned} \tag{46}$$

and $\xi_l(x) = (1/T^2) \sum_{t,j=1}^T \eta_{t,j,l}$. Then $|\eta_{t,j,l}| \leq 2\kappa^3$ and

$$\begin{aligned} & \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\| \\ &= \sup_{\|\beta\|_2 \leq 1} \left\| \left(L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x \right) \beta \right\|_\rho \\ &= \sup_{\|\beta\|_2 \leq 1} \left\| \frac{1}{T} \sum_{l=1}^T \beta_l \xi_l \right\|_\rho \end{aligned}$$

$$\begin{aligned} & \leq \frac{1}{T^3} \left(\int_X \sum_{l=1}^T \left(\sum_{t,j=1}^T \eta_{t,j,l} \right)^2 d\rho_X \right)^{1/2} \\ &= \frac{1}{T^3} \left(\int_X \sum_{t,j,w,\tau,l=1}^T \eta_{t,j,l}(x) \eta_{w,\tau,l}(x) d\rho_X \right)^{1/2}. \end{aligned} \tag{47}$$

Applying the same method as shown in the proof of Lemma 4.1 in [9], we can see that when all the indices t, j, w, τ , and l are pairwise different, there holds $\mathbb{E}_x(\eta_{t,j,l}(x)\eta_{w,\tau,l}(x)) = 0$; therefore

$$\begin{aligned} & \left(\mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} U_x - U_x S_x \widehat{U}_x S_x U_x\|^2 \right)^{1/2} \\ & \leq \frac{1}{T^3} \left(\mathbb{E}_x \left(\sum_{t,j,w,\tau,l=1}^T \eta_{t,j,l}(x) \eta_{w,\tau,l}(x) \right) \right)^{1/2} \\ & \leq \frac{2\kappa^3}{T^3} \sqrt{T^5 - T(T-1)(T-2)(T-3)(T-4)} \\ & \leq \frac{2\sqrt{10}\kappa^3}{T}. \end{aligned} \tag{48}$$

This together with (45) yields

$$\mathbb{E} \|II\|_\rho \leq \sqrt{2C_2 M} \left(\frac{2\sqrt{10}\kappa^3}{\lambda^{3/2} T^{1/2}} + \frac{\alpha C_1 \kappa^2 (\kappa + 2|K|_{C^*(X \times X)})}{\lambda^{3/2} T (1 - \alpha)} \right). \tag{49}$$

The term II is more involved; recall that

$$\begin{aligned} II &= \left(\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}} \right)^{-1} \left(U_x S_x \widehat{U}_x y - L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho \right) \\ &= \left(\lambda I + L_{\bar{K}, \bar{\rho}_X^{(T)}} \right)^{-1} \left\{ U_x S_x \widehat{U}_x y - L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho \right. \\ & \quad \left. + L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho \right\}. \end{aligned} \tag{50}$$

Hence

$$\begin{aligned} \mathbb{E} \|II\|_\rho & \leq \lambda^{-1} \left(\mathbb{E} \|U_x S_x \widehat{U}_x y - L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho\|_\rho \right. \\ & \quad \left. + \mathbb{E} \|L_{\bar{K}, \bar{\rho}_X^{(T)}} f_\rho - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho\|_\rho \right). \end{aligned} \tag{51}$$

If we define $\eta_{ij}(x) = y_t K(x_t, x_j) K(x, x_j)$ and

$$\begin{aligned} \xi_{ij}(x) &= \eta_{ij}(x) \\ & \quad - \iint_X K(x, u) K(v, u) f_\rho(v) d\rho_X^{(j)}(u) d\rho_X^{(t)}(v), \end{aligned} \tag{52}$$

$(t, j = 1, \dots, T),$

therefore

$$\begin{aligned} & \mathbb{E} \left\| U_x S_x \widehat{U}_x y - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho \right\|_\rho^2 \\ & \leq T^{-4} \sum_{t,j,w,\tau=1}^T \mathbb{E}_z \xi_{tj}^2(x) \xi_{w\tau}^2(x). \end{aligned} \tag{53}$$

If $t, j, w,$ and τ are pairwise distinct, then $\mathbb{E}_z(\xi_{tj}(x) \times \xi_{w\tau}(x)) = 0$. If $t = j$ or $w = \tau$,

$$\begin{aligned} \mathbb{E}_z \eta_{tt} & \neq \iint_X K(x, u) K(v, u) f_\rho(v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v), \\ \mathbb{E}_z \eta_{ww} & \neq \iint_X K(x, u) K(v, u) f_\rho(v) d\rho_X^{(w)}(u) d\rho_X^{(w)}(v). \end{aligned} \tag{54}$$

By the Cauchy-Schwartz inequality, for any $t, j, w, \tau = 1, \dots, T$,

$$\begin{aligned} \mathbb{E}_z \xi_{tj}(x) \xi_{w\tau}(x) & \leq \left(\mathbb{E}_z \xi_{tj}^2(x) \right)^{1/2} \left(\mathbb{E}_z \xi_{w\tau}^2(x) \right)^{1/2} \\ & \leq \max \{ \mathbb{E}_z \xi_{tj}^2(x), \mathbb{E}_z \xi_{w\tau}^2(x) \}. \end{aligned} \tag{55}$$

Hence we only need to give a bound for $\mathbb{E}_z \xi_{tj}^2$. Simple calculation shows

$$\begin{aligned} \mathbb{E}_z \xi_{tj}^2 & \leq \mathbb{E}_z \eta_{tj}^2 \\ & \leq \iint_X K^2(x, u) K^2(v, u) \\ & \quad \times \int_Y y^2 d\rho(y | v) d\rho_X^{(j)}(u) d\rho_X^{(t)}(v) \\ & \leq 2C_2 M^2 \kappa^4, \quad (t \neq j). \end{aligned} \tag{56}$$

By the same method, we know that

$$\begin{aligned} \mathbb{E}_z \xi_{tt}^2 & = \mathbb{E}_z \eta_{tt}^2 - 2\mathbb{E}_z \eta_{tt} \iint_X K(x, u) K(v, u) \\ & \quad \times \int_Y y d\rho(y | v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v) \\ & \quad + \left(\iint_X K(x, u) K(v, u) f_\rho(v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v) \right)^2 \\ & \leq \int_X K^2(v, v) K^2(x, v) \\ & \quad \times \int_Y y^2 d\rho(y | v) d\rho_X^{(t)}(v) + 2C_2 M^2 \kappa^4 \\ & \quad - 2 \int_X K(v, v) K(x, v) \int_Y y d\rho(y | v) d\rho_X^{(t)}(v) \\ & \quad \times \iint_X K(x, u) K(v, u) \\ & \quad \times \int_Y y d\rho(y | v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v) \\ & \leq 8C_2 M^2 \kappa^4. \end{aligned} \tag{57}$$

Applying the conclusion as shown in [9] and together with the above bound, we can see that

$$\begin{aligned} & T^{-4} \sum_{t,j,w,\tau=1}^T \mathbb{E}_z \xi_{tj}^2(x) \xi_{w\tau}^2(x) \\ & \leq 8T^{-4} (T^4 - T(T-1)(T-2)(T-3)) C_2 M^2 \kappa^4 \\ & \leq \frac{48C_2 M^2 \kappa^4}{T}. \end{aligned} \tag{58}$$

Hence

$$\begin{aligned} \mathbb{E} \left\| U_x S_x \widehat{U}_x y - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho \right\|_\rho & \leq \frac{4\sqrt{3C_2} M \kappa^2}{\sqrt{T}}, \\ \left\| L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho \right\|_\rho \\ & \leq \left\| \int_X (\widehat{K}(x, v) - \widehat{K}(x, v)) f_\rho(v) d\widehat{\rho}_X^{(T)}(v) \right\|_\rho \\ & \leq \|f_\rho\|_\infty \left\| \int_X K(x, u) K(v, u) d(\widehat{\rho}_X^{(T)} - \rho_X)(u) \right\|_\infty \\ & \leq C_2 M \left\| \widehat{\rho}_X^{(T)} - \rho_X \right\|_{(C^*(X))^*} \|K(x, \cdot) K(v, \cdot)\|_{C^*(X)} \\ & \leq \frac{\alpha C_1 \kappa (\kappa + 2|K|_{C^*(X \times X)}) C_2 M}{T(1-\alpha)}. \end{aligned} \tag{59}$$

This yields

$$\begin{aligned} & \mathbb{E} \left\| L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho - L_{\widehat{K}, \widehat{\rho}_X^{(T)}} f_\rho \right\|_\rho \\ & \leq \frac{\alpha C_1 \kappa (\kappa + 2|K|_{C^*(X \times X)}) C_2 M}{T(1-\alpha)}; \end{aligned} \tag{60}$$

then

$$\mathbb{E} \|III\|_\rho \leq \frac{4\sqrt{3C_2} M \kappa^2}{\lambda \sqrt{T}} + \frac{\alpha C_1 \kappa (\kappa + 2|K|_{C^*(X \times X)}) C_2 M}{\lambda T(1-\alpha)}. \tag{61}$$

This together with (49) yields the conclusion. \square

Now we are in a position to give the proofs of Theorems 5 and 7.

Proof of Theorem 5. Theorem 11 ensures that

$$\mathbb{E} \|f_{z,\lambda} - f_{\lambda, \widehat{\rho}_X^{(T)}}\|_\rho \leq \frac{C'_\kappa}{\lambda^{3/2} T^{1/2}}. \tag{62}$$

For $1/2 < r \leq 3/2$, Proposition 10 tells that

$$\|f_{\lambda, \widehat{\rho}_X^{(T)}} - f_{\lambda, \rho_X}\|_{\widehat{K}} \leq \frac{C_3 \lambda^{r-3/2}}{T}, \tag{63}$$

and Proposition 8 shows that

$$\|f_{\lambda, \rho_X} - f_\rho\|_{\widehat{K}} \leq C_r \lambda^{r-1/2}, \tag{64}$$

since

$$\|f\|_\rho \leq \|f\|_\infty \leq \kappa \|f\|_{\bar{\mathcal{K}}}, \quad \forall f \in \mathcal{H}_{\bar{\mathcal{K}}}. \quad (65)$$

Combining all the bounds together and noting that $\lambda = T^{-\theta}$ with $0 < \theta < 1/3$, we can get the conclusion of Theorem 5 by taking $C'_\kappa = (C_3 + C_r)\kappa + C'_\kappa$. \square

Proof of Theorem 7. When the samples are drawn i.i.d. from measure ρ , then $f_{\lambda, \bar{\rho}_X^{(T)}} = f_{\lambda, \rho_X}$. Hence

$$\begin{aligned} \mathbb{E} \|f_{z, \lambda} - f_\rho\|_\rho &\leq \mathbb{E} \|f_{z, \lambda} - f_{\lambda, \rho_X}\|_\rho \\ &+ \mathbb{E} \|f_{\lambda, \rho_X} - f_\rho\|_\rho \leq \frac{C'_\kappa}{\lambda^{3/2} T^{1/2}} + C_q \lambda^q. \end{aligned} \quad (66)$$

Let $\lambda = T^{-\theta}$; then

$$\mathbb{E} \|f_{z, \lambda} - f_\rho\|_\rho \leq \bar{C}_\kappa T^{-\min\{1/2 - (3/2)\theta, q\theta\}}. \quad (67)$$

The conclusion follows by discussing the relationship between r and 1. \square

Acknowledgments

The author would like to thank Professor Hongwei Sun for useful discussions which have helped to improve the presentation of the paper. The work described in this paper is supported partially by National Natural Science Foundation of China (Grant no. 11001247) and Doctor Grants of Guangdong University of Business Studies (Grant no. 11BS11001).

References

- [1] S. Smale and D.-X. Zhou, "Online learning with Markov sampling," *Analysis and Applications*, vol. 7, no. 1, pp. 87–113, 2009.
- [2] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [3] F. Cucker and D.-X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge University Press, Cambridge, UK, 2007.
- [4] S. Smale and D.-X. Zhou, "Learning theory estimates via integral operators and their approximations," *Constructive Approximation*, vol. 26, no. 2, pp. 153–172, 2007.
- [5] Q. Wu, Y. Ying, and D.-X. Zhou, "Learning rates of least-square regularized regression," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 171–192, 2006.
- [6] H. Sun and Q. Wu, "Indefinite kernel network with dependent sampling," *Analysis and Applications*. Accepted.
- [7] Q. Wu and D.-X. Zhou, "Learning with sample dependent hypothesis spaces," *Computers and Mathematics with Applications*, vol. 56, no. 11, pp. 2896–2907, 2008.
- [8] Q. Wu, "Regularization networks with indefinite kernels," *Journal of Approximation Theory*, vol. 166, pp. 1–18, 2013.
- [9] H. Sun and Q. Wu, "Least square regression with indefinite kernels and coefficient regularization," *Applied and Computational Harmonic Analysis*, vol. 30, no. 1, pp. 96–109, 2011.
- [10] L. Shi, "Learning theory estimate for coefficient-based regularized regression," *Applied and Computational Harmonic Analysis*, vol. 34, no. 2, pp. 252–265, 2013.
- [11] H. Sun and Q. Guo, "Coefficient regularized regression with non-iid sampling," *International Journal of Computer Mathematics*, vol. 88, no. 15, pp. 3113–3124, 2011.
- [12] J. B. Conway, *A Course in Operator Theory*, American Mathematical Society, 2000.
- [13] C. Wang and D.-X. Zhou, "Optimal learning rates for least squares regularized regression with unbounded sampling," *Journal of Complexity*, vol. 27, no. 1, pp. 55–67, 2011.
- [14] A. Caponnetto and E. De Vito, "Optimal rates for the regularized least-squares algorithm," *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 331–368, 2007.
- [15] E. De Vito, A. Caponnetto, and L. Rosasco, "Model selection for regularized least-squares algorithm in learning theory," *Foundations of Computational Mathematics*, vol. 5, no. 1, pp. 59–85, 2005.
- [16] S. Mendelson and J. Neeman, "Regularization in kernel learning," *The Annals of Statistics*, vol. 38, no. 1, pp. 526–565, 2010.
- [17] I. Steinwart, D. Hush, and C. Scovel, "Optimal rates for regularized least-squares regression," in *Proceedings of the 22nd Annual Conference on Learning Theory*, pp. 79–93, 2009.
- [18] S. Smale and D.-X. Zhou, "Shannon sampling. II. Connections to learning theory," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 285–302, 2005.
- [19] D.-X. Zhou, "Capacity of reproducing kernel spaces in learning theory," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1743–1752, 2003.