Least Square Regression with Non-Identical Unbounded Sampling and Coefficient Regularization

Xinxin Chang[1], Lu Luo[2]

[1]Heibei University of Technology, School of Science, 5340, Xiping Road, Beichen District, Tianj, China

[2]Heibei University of Technology, School of Science, 5340, Xiping Road, Beichen District, Tianjin, China

Abstract: ***In this paper, we consider 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. And except for continuity and boundedness, the kernel function is not necessary to satisfy any further regularity conditions. This leads to additional difficulty. By introducing a suitable reproducing kernel Hilbert space (RKHS) and a suitable intermediate integral operator, and by the error decomposition procedure the sample error is divided into two parts. We deduce the error bound .Last we yield satisfactory results by proper choice of the regularization parameter.***

Keywords: Indefinite Kernel; Coefficient Regularization; Least Square Regression; Integral Operator; Unbounded Hypothesis; Learning Rates

1. Introduction

The aim of this paper is to study coefficient-based least squares regression with indefinite kernels from non-identical unbounded sampling processes.

Let (X,d) be a compact metric space (input space), ρ be a probability distribution on $Z = X \times Y$ with $Y = R$ (out space). $\rho(y/x)$ is the conditional distribution according to ρ .The generalization error for function : $f : X \to Y$ is defined as

$$\varepsilon_\rho(f) = \int_Z (y - f(x))^2 d\rho$$

The regression function which minimizes the generalization error is given by

$$f_\rho(x) = \int_Y y d(y/x)$$

In most regression learning, the distribution $\rho(x,y)$ is unknown and what one can known is a set of samples $z = \{z_t\}_{t=1}^T = \{(x_t, y_t)\}_{t=1}^T \in Z^T$ are available. The aim of regression learning is to find a good estimator that describes the relationship between the input data x and output data y best through random sampling. This is ill-posed problem and the regularization technique is needed. The well-known regularized least square regression algorithm is conducted by a scheme in a reproducing kernel Hilbert space(RKHS)[1]associated with a Mercer kernel $K : X \times X \to R$,which is defined to be a continuous, symmetric, and positive semi-definite (p.s.d) function. RKHS 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 $k = \sup_{x,y \in X} |K(x,y)| < \infty$; then the regularized regression problem is given by

$$f_{z,\lambda} = \arg\min_{f \in Hk} \left\{ \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 + \lambda \|f\|_K^2 \right\}$$

It has been well understood due to lots of the literature ([2,3]).

In recent years, coefficient based regularization kernel network (CRKN) attract more attentions:

$$f_{z,\lambda} = \arg\min_{f \in Hk,x} \left\{ \frac{1}{m} \sum_{i=1}^m (f_\alpha(x_i) - y_i)^2 + \lambda \Omega_Z(f) \right\}, \lambda > 0$$

Where $\alpha = (\alpha_1, \ldots, \alpha_m) \in R^m$ and $f_\alpha = \sum_{i=1}^m \alpha_i K(x, x_i)$. The Penalty $\Omega_Z(f)$ is imposed on the coefficients of function $f \in \mathrm{H}_{K,x}$. In this setting, the hypothesis space H_k is replaced by a finite dimension function space:

$$\mathrm{H}_{K,x} = \left\{ f_\alpha(x) = \sum_{i=1}^m \alpha_i K(x, x_i) : \alpha = (\alpha_1, \ldots, \alpha_m) \in R^m, m \in \mathrm{N} \right\}$$

The application of $\mathrm{H}_{K,x}$ and the coefficient regularization was first introduced by Vapnik [4] to design programming support vector machines. And it has some advantages what we can see in [5].

In this article, we consider the general kernel, i.e. $K : X \times X \to R$ is continuous and bounded function. This kind of kernel scheme has been studied due to a lot of literature([6-8]).The learning algorithm we are interested in this paper takes the following form:

$$f_{Z,\lambda} = f_{\alpha Z},$$

$$where \alpha_Z = \arg\min_{\alpha \in R^m} \left\{ \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 + \lambda m \sum_{i=1}^m \alpha_i^2 \right\}, \lambda > 0 \tag{1.1}$$

By using the integral operator technique from [9],in [7] Wu gave the capacity independent estimate for the convergence rate for the indefinite kernels. Sun and Guo conducted error analysis for the Mercer kernels but uniform bounded non-i.i.d. sampling[5]. In this paper, we study learning algorithm (1.1) by non-identical unbounded sampling processes with indefinite kernels.

2. Assumption and Main Results

We study coefficient-based least squares regression with indefinite kernels from non-identical unbounded sampling processes. In our setting, a sample $z=\{z_t=(x_t,y_t)\}_{t=1}^{T}\subset Z^T$ is drawn independently from different Borel probability measures $\rho^t(t=1,\ldots,T),\rho^t(\cdot/x)=\rho(\cdot/x)$. 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 Holder space $C^s(X)$. Here the Holder space $C^s(X)(0\le s\le 1)$ is defined as the space of all continuous functions on X with the following norm finite[5]:

$$\left\|f\right\|_{C^S(x)}=\left\|f\right\|_{\infty}+\left|f\right|_{C^S(x)}$$

Where

$$\left|f\right|_{C^S(x)}:=\sup_{x\ne y\in X}\frac{|f(x)-f(y)|}{(d(x,y))^s}.$$

Definition 1. Let $0\le s\le 1$;we say that the sequence $\{\rho_X^{(t)}\}$ convergence exponentially fast in $(C^s(X))^*$ to a probability measure ρ_X on X or convergence exponentially in short if there exist $C_1>0$ and $0<\alpha<1$ such that

$$\left\|\rho_X^{(t)}-\rho_X\right\|_{(C^S(X))^*}\le C_1\alpha^t,\forall t\in \mathrm{N}. \tag{2.1}$$

By the definition of the dual space $(C^s(X))^*$,the decay condition (2.1) can be expressed as

$$\left|\int_X f(x)d\rho_X^{(t)}-\int_X f(x)d\rho_X\right|\le C_1\alpha^t\left\|f\right\|_{C^S(X)},\forall f\in C^s(X),t\in \mathrm{N}.$$

For the bounded indefinite kernel K(x,y),we consider the Mercer kernel

$$\tilde{K}(u,v)=\int_X K(u,x)K(v,x)d\rho_X(x).$$

For more introduction about learning with indefinite kernels, please see [6-8].

For all $x\in X$, we can define $L_{K,\rho X}f(x)=\int_X K(x,u)f(u)d\rho_X(\mu)$ 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, $L_{K,\rho X}$ and its adjoint $L^*_{X,\rho X}$ are both compact operators. Hence $L_{\tilde{K},\rho X}=L_{K,\rho X}L^*_{K,\rho X}, L_{\tilde{K},\rho_X^{(t)}}=L_{K,\rho_X^{(t)}}L^*_{K,\rho_X^{(t)}}$.

If K is a Mercer kernel, from [10] we know that H_k is in the range of $L_K^{\frac{1}{2}}$. For an indefinite kernel K, recall $L_{\tilde{K},\rho X}=L_{K,\rho X}L^*_{K,\rho X}$ .Based on the polar decomposition of compact operators ([11]).

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

$$T=\Gamma A$$

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

We immediately have the following proposition[12].

Proposition3. Consider $\mathrm{H}_{\tilde{K}}$ as a subspace of $L^2_{\rho X}$;then $L_K^*=UL_{\tilde{K}}^{1/2}$ and $L_K=L_K^{1/2}U^*$,where U is a partial isometry on $L^2_{\rho X}$ with U^*U being the orthogonal projection onto $\overline{\mathrm{H}}_{\tilde{K}}$.

We use the RKHS $\mathrm{H}_{\tilde{K}}$ to approximate $f_{\rho,}$ hence define

$$f_{\lambda,\rho X}=\arg\min_{f\in \mathrm{H}_{\tilde{K}}}\left\{\int_Z(f(x)-f_\rho(x))^2d\rho(x,y)+\lambda\left\|f\right\|_{\tilde{K}}^2\right\} \tag{2.2}$$

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

$$f_{\lambda,\bar{\rho}_X^{(T)}}=(\lambda I+L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}L_{\tilde{K},\bar{\rho}_X^{(T)}}f_\rho \tag{2.3}$$

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:

$$\left\|f_{z,\lambda}-f_\rho\right\|_\rho=\left\|f_{z,\lambda}-f_{\lambda,\bar{\rho}_X^{(T)}}\right\|_\rho+\left\|f_{\lambda,\bar{\rho}_X^{(T)}}-f_{\lambda,\rho X}\right\|_\rho$$
$$+\left\|f_{\lambda,\rho X}-f_\rho\right\|_\rho$$

where the first term on the right hand side is sample error , the second term is measure error and the third one is regularization error.

We will conduct the error analysis in several steps. The first major contribution we make is on the sample error estimate; the main difficulty is the non-identical unbounded sampling of the samples; we overcome that by introducing a suitable intermediate operator. The second one we make is on the measure error estimate, there we get a sharp error bound by a new method.

In order to give the error analysis, we assume that the kernel $\tilde{K}$ satisfies the following kernel condition [5].

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

$$\left\|\tilde{K}_u-\tilde{K}_v\right\|_{\tilde{K}}\le k_s(d(u,v))^s. \tag{2.4}$$

Since sample Z is drawn from unbounded sampling processes, we will assume the following unbounded hypothesis[13]:

Unbounded hypothesis: There exists some constant $M>0$ such that

$$\int_Z y^2d\rho\le M^2 \tag{2.5}$$

Remark 1.Theoretical study of learning algorithms for regression is mostly based on the standard assumption that $|y|\le M$ almost surely for some constant M>0.In [14] the author consider a general setting satisfying the moment hypothesis:

Moment hypothesis: There exist constants $M>0$ and $C_2>0$ such that

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

One of the main purposes of this paper is to improve the moment hypothesis to unbounded hypothesis and consider coefficient regularization algorithms with non-identical sampling, and the other is to deduced the measure error by a new method (In section 3.2).

By a simple computation, it follows that

$$\int_X f_\rho^2(x) d\rho_X \le \int_X (\int_Y y d\rho(y|x)^2) d\rho_X \le \int_Z y^2 d\rho \le M^2$$

Therefore, the regression function f_ρ is square integral with respect to the marginal distribution ρ_X and the variance of ρ is finite, i.e., $f_\rho \in L^2_{\rho X}(X)$ and $\sigma^2 = \mathrm{E}(y - f_\rho(x))^2 < \infty$. Now, we can state our general results on learning rates for algorithm (2).

Theorem 5. Assume unbounded hypothesis condition (2.5); $\rho_X^{(t)}$ satisfies condition(2.1)and $\tilde{K}$ satisfies condition(2.4); $L^{-r}_{\tilde{K},\rho X} f_\rho \in L^2_{\rho X}$,for $1/2 < r \le 3/2$;take $\lambda = T^{-\theta}$ with $0<\theta<1/3$; then

$$E\left\|f_{z,\lambda} - f_\rho\right\|_\rho \le C_k^{"} T^{-\min\{1/2-(3/2)\theta,(r-1/2)\theta\}},$$

where $C_k^{"}$ is a constant depending on k, s and α ,but not on T or δ ,and will be given explicitly in Section 3.3.

Remark 6.The proof of Theorem 5 will be conducted in Section 3, where the error term is decomposed into three parts. In [5], 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 unbounded hypothesis condition (2.5); $\tilde{K}$ satisfies condition(2.4); $L^{-r}_{\tilde{K},\rho X} f_\rho \in L^2_{\rho X}$;

then if $0<r\le 1$,take $\lambda = T^{-1/(2r+3)}$; one see that

$$\left\|f_{z,\lambda} - f_\rho\right\|_\rho = o(T^{-r/(2r+3)}),$$

And if r>1,take $0<r\le 1$, $\lambda = T^{-1/5}$;we have

$$\left\|f_{z,\lambda} - f_\rho\right\|_\rho = o(T^{-1/5}).$$

Here we get the same learning rate as one in [8].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 $\left\|f_{\lambda,\rho X} - f_\rho\right\|$ .The error estimate for regularization error has been investigated in lots of the literature in learning theory ([9] and the references therein); we will omit the poof and quote it directly.

Proposition 8. Assume $L^{-r}_{\tilde{K},\rho X} f_\rho \in L^2_{\rho X}$ and r>0; the following bound for approximation error holds:

$$\left\|f_{\lambda,\rho X} - f_\rho\right\|_\rho \le C_q \lambda^q$$

where $C_q = (1+k^{2r-2})\left\|L^{-r}_{\tilde{K},\rho X} f_\rho\right\|_\rho, q = \min\{r,1\}$,and when $1/2 < r \le 3/2$,

$$\left\|f_{\lambda,\rho X} - f_\rho\right\|_{\tilde{K}} \le C_r \lambda^{r-1/2},$$

where $C_r = \left\|L^{-r}_{\tilde{K},\rho X} f_\rho\right\|_\rho$.

3.2. Estimate for the Measure Error

This subsection is devoted to the analysis of the term $\left\|f_{\lambda,\bar{\rho}_X^{(T)}} - f_{\lambda,\rho X}\right\|_{\tilde{K}}$ caused by the difference of measures, which we called measure error. The ideas of proof are from[5].Before giving the result, let us state two lemmas .

Firstly, we need the following lemma which can be proved by easy computation.

Lemma 9. For any $f, g \in C^s(X)$, we have

$$\|fg\|_{C^{(s)}(X)} \le \|f\|_{C^s(X)} \times \|g\|_\infty + \|f\|_\infty \times |g|_{C^s(X)}.$$

Lemma 10. Assume $\tilde{K}$ satisfies condition (12);then

$$\left\|L_{\tilde{K},\bar{\rho}_X^{(T)}} - L_{\tilde{K},\rho_X}\right\| \le k(k+2k_s)\left\|\bar{\rho}_X^{(T)} - \rho\right\|_{(C^s(X))^*}$$

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

$$\left\|(L_{\tilde{K},\bar{\rho}_X^{(I)}} - L_{\tilde{K},\rho X})h\right\|^2_{\tilde{K}}$$

$$= \int_X h(u)\{\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)(v)\}d(\bar{\rho}_X^{(T)} - \rho_X)(u)$$

$$\le \left\|h(u)\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right\|_{C^s(X)} \times \left\|\bar{\rho}_X^{(T)} - \rho_X\right\|_{(C^s(X))^*}$$

$$\le [\|h\|_{C^s(X)} \times \left\|\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right\|_\infty + \|h\|_\infty \times \left|\int_X h(v)\tilde{K}(u,v) \times d(\bar{\rho}_X^{(T)} - \rho_X)\right|_{C^s(X)}] \times \left\|\bar{\rho}_X^{(T)} - \rho_X\right\|_{(C^s(x))^*}. \quad (3.1)$$

Denote

$$I = \left\|\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right\|_\infty,$$
$$II = \left|\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right|_{C^s(X)}.$$

Now we need to estimate I and II respectively. For the term I,it is easy to see that

$$I = \max_{x\in X}\left|\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right| \le \max_{x\in X}\left\|h(\cdot)\tilde{K}(u,\cdot)\right\|_{C^s(X)} \times \left\|\bar{\rho}_X^{(T)} - \rho_X\right\|_{(C^s(X))^*}$$

$$\le \max_{x\in X}[\|h\|_{C^s(X)} \times k^2 + \|h\|_\infty \times \left|\tilde{K}(u,\cdot)\right|_{C^s(X)}] \times \left\|\bar{\rho}_X^{(T)} - \rho_X\right\|_{(C^s(X))^*}$$

$$\le [k^2\|h\|_{C^s(X)} + kk_s\|h\|_\infty] \times \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(x))^*} \qquad (3.2)$$

In order to estimate II, we consider

$$\left|\int_X h(v)[\tilde{K}(u_2,v) - \tilde{K}(u_1,v)]d(\bar{\rho}_X^{(T)} - \rho_X)\right|$$
$$\le \|h(v)[\tilde{K}(u_2,v) - \tilde{K}(u_1,v)]\|_{C^s(X)} \times \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}$$
$$\le [\|h\|_{C^s(X)} \times \|\tilde{K}(u_2,v) - \tilde{K}(u_1,v)\|_\infty + \|h\|_\infty |\tilde{K}(u_2,t) - \tilde{K}(u_1,t)|_{C^s(X)}]$$
$$\times \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}. \qquad (3.3)$$

Since

$$|[\tilde{K}(u_2,v_2) - \tilde{K}(u_1,v_2)] - [\tilde{K}(u_2,v_1) - \tilde{K}(u_1,v_1)]| \le k_s^2 d^s(u_1,u_2) d^s(v_1,v_2),$$

then

$$|\tilde{K}(u_2,v) - \tilde{K}(u_1,v)|_{C^s(X)} \le k_s^2 d^s(u_1,u_2), |\tilde{K}(u_2,v) - \tilde{K}(u_1,v)|$$
$$\le k \cdot k_s d^s(u_1,u_2).$$

Taking these estimates into(3.3),we have

$$II = \left|\int_X h(v)\tilde{K}(u,v)d(\bar{\rho}_X^{(T)} - \rho_X)\right|_{C^s(X)}$$
$$\le (kk_s\|h\|_{C^s(X)} + k_s^2\|h\|_\infty) \cdot \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}. \qquad (3.4)$$

Putting the estimates (3.2)and(3.4)into(3.1)

$$\left\|(L_{\tilde{K},\bar{\rho}_X^{(I)}} - L_{\tilde{K},\rho X})h\right\|_{\tilde{K}}^2$$
$$\le \{\|h\|_{C^s(X)}[k^2\|h\|_{C^s(X)} + kk_s\|h\|_\infty] + \|h\|_\infty[kk_s\|h\|_{C^s(X)} +$$
$$k_s^2\|h\|_\infty]\} \times \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}^2$$
$$= (k\|h\|_{C^s(X)} + k_s\|h\|_\infty)^2 \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}^2$$

When condition (2.4)is satisfied, it was proved in[5]that $H_{\tilde{K}}$ is included in $C^s(X)$ with the inclusion bounded

$$\|h\|_{C^s(X)} \le (k+k_s)\|h\|_{\tilde{K}}, \forall h \in H_{\tilde{K}}.$$

Then we obtain

$$\left\|(L_{\tilde{K},\bar{\rho}_X^{(T)}} - L_{\tilde{K},\rho X})h\right\|_{\tilde{K}} \le k(k+2k_s)\|h\|_{\tilde{K}} \|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*}.$$

This completes the proof.

Proposition 11. Assume $L_{\tilde{K},\rho X}^{-r} f_\rho \in L_{\rho X}^2$ for some $1/2 < r \le 3/2$; $\tilde{K}$ satisfies condition (2.4); the following bound for measure error holds:

$$\left\|f_{\lambda,\bar{\rho}_X^{(T)}} - f_{\lambda,\rho X}\right\|_{\tilde{K}} \le \frac{C_3\lambda^{r-3/2}}{T}$$

where $C_k = k(k+2k_s)$, and $C_3 = C_1C_kC_r\alpha/(1-\alpha)$.

Proof. From(2.2), simple calculation shows that $f_{\lambda,\rho X} = (\lambda I + L_{\tilde{K},\rho X})^{-1} L_{\tilde{K},\rho X} f_\rho$. Recalling (2.3), we can see that

$$\left\|f_{\lambda,\bar{\rho}_X^T} - f_{\lambda,\rho X}\right\|_{\tilde{K}}$$
$$= \left\|(\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}\{(L_{\tilde{K},\bar{\rho}_X^{(T)}} - L_{\tilde{K},\rho X}) f_\rho + (L_{\tilde{K},\rho X}\right.$$
$$\left. - L_{\tilde{K},\bar{\rho}_X^{(T)}}) f_{\lambda,\rho X}\}\right\|_{\tilde{K}}$$
$$= \left\|(\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(L_{\tilde{K},\bar{\rho}_X^{(T)}} - L_{\tilde{K},\rho X})(f_\rho - f_{\lambda,\rho X})\right\|_{\tilde{K}}$$
$$\le \frac{1}{\lambda}\left\|(L_{\tilde{K},\bar{\rho}_X^{(T)}} - L_{\tilde{K},\rho X})(f_\rho - f_{\lambda,\rho X})\right\|_{\tilde{K}}.$$

Applying Lemma 10 to the case $h = f_\rho - f_{\lambda,\rho X}$,we get

$$\left\|f_{\lambda,\bar{\rho}_X^{(T)}} - f_{\lambda,\rho X}\right\|_{\tilde{K}}$$
$$\le \frac{1}{\lambda} k(k+2k_s)\|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*} \|f_{\lambda,\rho X} - f_\rho\|_{\tilde{K}}.$$

By the definition of $\bar{\rho}_X^{(T)}$ and noticing (2.1),we can see

$$\|\bar{\rho}_X^{(T)} - \rho_X\|_{C^s(X)^*} = \frac{1}{T}\sum_{t=1}^T \|\rho_X^{(t)} - \rho_X\|_{(C^s(X))^*} \le \frac{C_1\alpha}{T(1-\alpha)}.$$

This in connection with Proposition 8 yields the conclusion.

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 notions. Let $C(X)$ be the space of bounded continuous functions on X with supremum norm $\|\cdot\|_\infty$ .Define the sampling operator $S = S_x$ associated to the sampling points $x = \{x_1,\ldots,x_T\}$ as follows:

$$S : C(X) \to R^T$$
$$f \to (f(x_1),\ldots,f(x_T)).$$

Let U and U^* be operators from R^T to $C(X)$ defined as,for $\alpha = (\alpha_1,\ldots,\alpha_T)$,

$$U\alpha = \frac{1}{T}\sum_{i=1}^T \alpha_i K(x,x_i)$$
$$U^*\alpha = \frac{1}{T}\sum_{i=1}^T \alpha_i K(x_i,x)$$

It is easy to see that both U and U^* are bounded operator. From[8],we have

$$f_{z,\lambda} = U(\lambda I + SU^*SU)^{-1} SU^* y.$$

where **y**= $(y_1,\ldots,y_T)'$.

Employing the method as shown in[8].Let $g_{z,\lambda} = (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1} USU^* y$,

than we can decompose the sample error into two parts:

$$f_{z,\lambda} - f_{\lambda,\bar{\rho}_X^{(T)}} = (f_{z,\lambda} - g_{z,\lambda}) + (g_{z,\lambda} - f_{\lambda,\bar{\rho}_X^{(T)}})$$

$$= \{U(\lambda I + SU^*SU)^{-1} SU^* y - (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1} USU^* y\} +$$
$$\{(\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1} USU^* y - (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1} L_{\tilde{K},\bar{\rho}_X^{(T)}} f_\rho\}$$
$$= I + II$$

Now, let us estimate 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 $L_{\hat{K},\bar{\rho}_X^{(T)}}$, which plays an intermediate role in the estimates, and the definition of it will be given later.

Theorem 12. Let $f_{z,\lambda}$ be given by(1.1),assume unbounded hypothesis condition(2.5),and the marginal distribution sequence $\rho_X^{(t)}, t\in \mathbb{N}$,satisfies condition(2.1);then

$$E\left\| f_{z,\lambda} - f_{\lambda,\bar{\rho}_X^{(T)}} \right\|_\rho \le \frac{C_k^{'}}{\lambda^{3/2}T^{1/2}},$$

Where

$$C_4 = \alpha C_1 k M^2 (k+2|k|_{C^s(X\times X)})(k+1),$$

$$C_k^{'} = 2\sqrt{6}Mk^2 + 2\sqrt{10}Mk^3 + C_4/(1-\alpha).$$

Proof. We will estimate I and II, respectively.

$$I = (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})\times U(\lambda I + SU^*SU)^{-1}SU^*y -$$
$$(\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}U(\lambda I + SU^*SU)\times(\lambda I + SU^*SU)^{-1}SU^*y$$

$$= (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(L_{\tilde{K},\bar{\rho}_X^{(T)}}U - USU^*SU)\times(\lambda I + SU^*SU)^{-1}SU^*y$$

$$= (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(L_{\tilde{K},\bar{\rho}_X^{(T)}}U - USU^*SU)T\alpha_{z,\lambda}.$$

Where $\alpha_{z,\lambda} = \frac{1}{T}(\lambda I + SU^*SU)^{-1}SU^*y$.

Then

$$\|I\|_\rho \le \lambda^{-1}T\left\| L_{\hat{K},\bar{\rho}_X^{(T)}} - USU^*SU \right\| \|\alpha_{z,\lambda}\|_2 + \lambda^{-1}T\times$$
$$\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\| \|\alpha_{z,\lambda}\|_2$$

where $\hat{K}(x,t) = \int_X K(x,v)K(t,v)d\bar{\rho}_X^{(T)}(v)$ and $\|\alpha_{z,\lambda}\|_2$ is the l^2 norm on R^T, for $\|\alpha_{z,\lambda}\|_2$;noticing

$$\alpha_{z,\lambda} = \arg\min_{\alpha\in R^T}\{\frac{1}{T}\sum_{i=1}^{T}(f_\alpha(x_i)-y_i)^2 + \lambda T\sum_{i=1}^{T}\alpha_i^2\},$$

We can have

$$\lambda T\|\alpha_{z,\lambda}\|_2^2 \le \frac{1}{T}\sum_{i=1}^{T}y_i^2 .$$

This means

$$E\|I\|_\rho \le \lambda^{-1}T(E\|\alpha_{z,\lambda}\|_2^2)^{1/2}\{(E\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\|^2)^{1/2} +$$
$$(E\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\|^2)^{1/2}\}$$
$$\le \lambda^{-3/2}\sqrt{T}(E(\frac{1}{T}\sum_{i=1}^{T}y_i^2))^{1/2}\{(E\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\|^2)^{1/2} +$$
$$(E\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\|^2)^{1/2}\}$$
$$\le \lambda^{-3/2}\sqrt{T}M\{(E\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\|^2)^{1/2} + \qquad .$$
$$(E\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\|^2)^{1/2}\}$$

According to the definition of U,for any $\alpha\in R^T$, $\|U\alpha\|_\rho \le (k/\sqrt{T})\|\alpha\|_2$; this implies that $\|U\| \le k/\sqrt{T}$. Therefor

$$\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\|$$
$$= \sup_{\|\alpha\|_2\le 1}\left\| \frac{1}{T}\sum_{i=1}^{T}\alpha_i\int_X(\tilde{K}(x,u)-\hat{K}(x,u))\times K(u,x_i)d\bar{\rho}_X^{(T)}(u) \right\|_\rho$$
$$\le \sup_{\|\alpha\|_2\le 1}\frac{1}{T}\sum_{i=1}^{T}|\alpha_i|\|(\int_X K^2(u,x_i)d\bar{\rho}_X^{(T)}(u))^{1/2}\times$$
$$(\int_X[\tilde{K}(x,u)-\hat{K}(x,u)]^2\times d\bar{\rho}_X^{(T)}(u))^{1/2}\|_\rho$$
$$\le \frac{k}{\sqrt{T}}\left\| \int_X K(x,v)K(u,v)d(\bar{\rho}_X^{(T)}-\rho_X)(v) \right\|_\infty$$
$$\le \frac{k}{\sqrt{T}}\left\| \bar{\rho}_X^{(T)}-\rho_X \right\|_{(C^s(X))^*}\|K(x,\cdot)K(u,\cdot)\|_{C^s(X)}$$
$$\le \frac{\alpha C_1 k^2(k+2|K|_{C^s(X\times X)})}{T^{3/2}(1-\alpha)}.$$

Hence

$$(E\left\| L_{\tilde{K},\bar{\rho}_X^{(T)}}U - L_{\hat{K},\bar{\rho}_X^{(T)}}U \right\|^2)^{1/2} \le \frac{\alpha C_1 k^2(k+2|K|_{C^s(X\times X)})}{T^{3/2}(1-\alpha)}. \tag{3.5}$$

For the term $\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\|$, let

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

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

$$\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\| = \sup_{\|\alpha\|_2\le 1}\left\| (L_{\hat{K},\bar{\rho}_X^{(T)}} - USU^*SU)\alpha \right\|_\rho$$
$$= \sup_{\|\alpha\|_2\le 1}\left\| \frac{1}{T}\sum_{l=1}^{T}\alpha_l\xi_l \right\|_\rho$$
$$\le \frac{1}{T^3}(\int_X\sum_{l=1}^{T}(\sum_{t,j=1}^{T}\eta_{t,j,l})^2)^{1/2}$$
$$= \frac{1}{T^3}(\int_X\sum_{t,j,w,\tau,l}^{T}\eta_{t,j,l}(x)\eta_{w,\tau,l}(x)d\rho_X)^{1/2}$$

Applying the same method as shown in the proof of Lemma 4.1 in [8],we can see that when all the indices t,j,w,τ and l are pairwise different, there holds $(E_X(\eta_{t,j,l}(x)\eta_{w,\tau,l}(x))=0$; and we have

$$(E\left\| L_{\hat{K},\bar{\rho}_X^{(T)}}U - USU^*SU \right\|^2)^{1/2} \le \frac{2\sqrt{10}k^3}{T}$$

This together with(4.5)yields

$$E\|I\|_\rho \le M(\frac{2\sqrt{10}k^3}{\lambda^{3/2}T^{1/2}} + \frac{\alpha C_1 k^2(k+2|K|_{C^s(X\times X)})}{\lambda^{3/2}T(1-\alpha)}). \tag{3.6}$$

The term II is more involved;recall that

$$II = (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(USU^*y - L_{\tilde{K},\bar{\rho}_X^{(T)}}f_\rho)$$

$$= (\lambda I + L_{\tilde{K},\bar{\rho}_X^{(T)}})^{-1}(USU^*y - L_{\hat{K},\bar{\rho}_X^{(T)}}f_\rho + L_{\hat{K},\bar{\rho}_X^{(T)}}f_\rho - L_{\tilde{K},\bar{\rho}_X^{(T)}}f_\rho)$$

Hence

$$E\|II\|_\rho \le \lambda^{-1}(E\|USU^* y - L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho + E\|L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho - L_{\tilde{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho)$$

Firstly, if we define $\eta_{t,j}(x) = y_t K(x_t, x_j) K(x, x_j)$ and $\xi_{t,j}(x) = \eta_{t,j}(x) - \iint_X K(x,u)K(v,u) f_\rho(v) d\rho_X^{(j)}(u) d\rho_X^{(t)}(v), (t, j = 1,...,T)$,

there

$$E\|USU^* y - L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho^2 \le T^{-4} \sum_{t,j,w,\tau=1}^{T} E_Z \xi_{t,j}(x) \xi_{w,\tau}(x).$$

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

$$E_Z \eta_{tt} \ne \iint_X K(x,u)K(v,u) f_\rho(v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v)$$

$$E_Z \eta_{ww} \ne \iint_X K(x,u)K(v,u) f_\rho(v) d\rho_X^{(w)}(u) d\rho_X^{(w)}(v)$$

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

$$E_Z \xi_{tj}(x)\xi_{w\tau}(x) \le (E_Z \xi_{tj}{}^2(x))^{1/2} (E_Z \xi_{w\tau}{}^2(x))^{1/2}$$
$$\le \max\{E_Z \xi_{tj}{}^2(x), E_Z \xi_{w\tau}{}^2(x)\}.$$

Hence we only need to give a bound for $E_Z \xi_{tj}{}^2(x)$. Simple calculation shows

$$E_Z \xi_{tj}^2 \le E_Z \eta_{tj}^2$$
$$\le \iint_X K^2(x,u)K^2(v,u) \times \int_Y y^2 d\rho(y/v) d\rho_X^{(j)}(u) d\rho_X^{(t)}(v)$$
$$\le M^2 k^4, (t \ne j)$$

By the same, we known that

$$E_Z \xi_{tt}^2$$
$$= E_Z \eta_{tt} - 2E_Z \eta_{tt} \iint_X K(x,u)K(v,u) \times \int_Y y d\rho(y/v) d\rho_X^{(t)} d\rho_X^{(t)}$$
$$+ (\iint_X K(x,u)K(v,u) f_\rho(v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v))^2$$
$$\le \int_X K^2(v,v)K^2(x,v) \times \int_Y y^2 d(y/v) d\rho_X^{(t)}(v) + M^2 k^4 -$$
$$2\int_X K(v,v)K(x,v) \times \int_Y y d\rho(y/v) d\rho_X^{(t)}(v) \iint_X K(x,u)K(v,u) \int_Y y$$
$$d\rho(y/v) d\rho_X^{(t)}(u) d\rho_X^{(t)}(v)$$
$$\le 4M^2 k^4.$$

Applying the conclusion as shown in[8]and together with the above bound, we can see that

$$T^{-1} \sum_{t,j,w,\tau=1}^{T} E_Z \xi_{tj}(x) \xi_{w\tau}(x)$$
$$\le 4T^{-4}(T^4 - T(T-1)(T-2)(T-3)) M^2 k^4.$$
$$\le \frac{24M^2k^4}{T}.$$

Hence

$$E\|USU^* y - L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho \le \frac{2\sqrt{6}Mk^2}{\sqrt{T}},$$

While

$$\|L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho - L_{\tilde{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho^2$$
$$\le \|\int_X (\hat{K}(x,v) - \tilde{K}(x,v)) f_\rho(v) d\bar{\rho}_X^{(T)}\|_\rho^2$$
$$\le \|f_\rho\|_\infty^2 \|\int_X K(x,u)K(v,u) d(\bar{\rho}_X^{(T)} - \rho_X)(u)\|_\infty^2$$
$$\le M^2 \|\bar{\rho}_X^{(T)} - \rho_X\|_{C^s(X)}^2 \|K(x,\cdot)K(v,\cdot)\|_{C^s(X)}^2$$

This yields

$$E\|L_{\hat{K},\bar{\rho}_X^{(T)}} f_\rho - L_{\tilde{K},\bar{\rho}_X^{(T)}} f_\rho\|_\rho$$
$$\le M\|\bar{\rho}_X^{(T)} - \rho_X\|_{(C^s(X))^*} \|K(x,\cdot)K(v,\cdot)\|_{C^s(X)}$$
$$\le \frac{\alpha C_1 k(k + 2|K|_{C^s(X\times X)})M}{T(1-\alpha)}.$$

Then

$$E\|II\|_\rho \le \frac{2\sqrt{6}Mk^2}{\lambda\sqrt{T}} + \frac{\alpha C_1 k(k+2|K|_{C^s(X\times X)})M}{\lambda T(1-\alpha)}.$$

This together with(3.6)yields the conclusion.

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

Proof of Theorem 5.Theorem 12 ensures that $E\|f_{z,\lambda} - f_{\lambda,\bar{\rho}_X^{(T)}}\|_\rho \le \frac{C_k'}{\lambda^{3/2} T^{1/2}}.$

For $1/2 < r \le 3/2$, Proposition11 tells that

$$\|f_{\lambda,\bar{\rho}_X^{(T)}} - f_{\lambda,\rho X}\|_{\tilde{K}} \le \frac{C_3 \lambda^{r-3/2}}{T},$$

and Proposition 8 shows that

$$\|f_{\lambda,\rho X} - f_\rho\|_{\tilde{K}} \le C_r \lambda^{r-1/2},$$

Since

$$\|f\|_\rho \le \|f\|_\infty \le k\|f\|_{\tilde{K}}, \forall f \in \mathrm{H}_{\tilde{K}}.$$

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

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

$$E\|f_{z,\lambda} - f_\rho\|_\rho \le E\|f_{z,\lambda} - f_{\lambda,\rho X}\|_\rho + E\|f_{\lambda,\rho X} - f_\rho\|_\rho$$
$$\le \frac{C_k^{'}}{\lambda^{3/2} T^{1/2}} + C_q \lambda^q$$

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

$$E\|f_{z,\lambda} - f_\rho\|_\rho \le \tilde{C}_k T^{-\min\{1/2-(3/2)\theta, q\theta\}}$$

The conclusion follows by discussion the relationship between r and 1.

4.Discussion

In this paper, we considered the learning performance of coefficient regularized least square regression based on non-identical and unbounded sampling. The only conditions we impose on the kernel function are the continuity and bounded. Capacity independent error bounds are derived by the integral operator technique.

From our results we can make the following observations.

- First, we deduced the error bound by deducing the measure error. And the measure error is much smaller than the one in[15].
- Second, we have reduced the Moment Hypothesis, it made we get satisfactory results under mild conditions.

References

[1] N.Aronszajn, "Theory of reproducing kernels", Transaction of the American Mathematical Society,LXVIII(3), pp.337-404,1969.(journal style).
[2] Q.Wu,Y.Ying,D-X.Zhou,"Learning rates of least-square regularized regression",Foundations of Computational Mathematics,VI(2),pp.171-192,2006.(journal style).
[3] Y-L.Xu , D-R.Chen , H-X.Li ,L.Liu"Least Square Regularized Regression in Sum Space",IEEE Transactions on Neural Networks & Learning Systems, XXIV(4),pp.355,2013.(journal style).
[4] Vapink.V."Statistical learning theory",Wiley,New York,XLI(4),pp.3185-3185,1998.(journal style).
[5] H.Sun ,Q.Guo,"Coefficient regularized regression with non-iid sampling",International Journal of Computer Mathematics,LXXXVIII(15),pp.3113-3124,2011.(journal style).
[6] H.Sun ,Q.Wu, "Indefinite kernel networks with dependent sampling", Analysis and Application Accepted.XXI(5), pp.880-646,2013.(journal style).
[7] Q.Wu, "Regularization networks with indefinite kernels", Journal of Approximation Theory, CLXVI (3),pp.1-18, 2013.(journal style).
[8] H.Sun ,Q.Wu, "Least square regression with indefinite kernels and coefficient regularization", Applied and Computation Harmonic Analysis,XXX(1),pp.96-109, 2011.(journal style).
[9] S.Smale ,D.-X.Zhou, "Learning theory estimates via integral operators and their approximations", Constructive Approximation, XXVI(2),pp.153-172, 2007.(journal style).
[10] F.Cucker,D.-X.Zhou,"Learning Theory:An Approximation Theory Viewpoint",Cambridge University Press,Cambridge,UK,2007.(journal style).
[11] J.B.Conway,"A Course in Operator Theory",American Mathematics Society,2000.(journal style).
[12] L.Shi,"learning theory estimates for coefficient-based regularized Regression",Applied Computational Harmonic Analysis,XXXIV(2),pp.252-265,2013.(journal style).
[13] Xiaorong ,H.Sun,"Coefficient Regularization with Unbounded Sampling",Journal of Computational Information System ,VIII(4),pp.1613-1621,2012.(journal style).
[14] J.Cai , C .Wang,"Coefficient-Based Regularize Regression with Indefinite Kernels by Unbounded Sampling",SciSin Math, XLIII(6)pp.613-624, 2013.(journal style).
[15] Cai,J."Coefficient-Based Regression with Non-Identical Unbounded Sampling",Abstract and Applied Analysis, MMXIII(39),pp.415-425,2013.(journal style).