Learning Rates of Regularized Regression with p-loss

SHENG BAOHUAI, YU WANGKE
Department of Mathematics,
Shaoxing College of Arts and Sciences
Shaoxing, Zhejiang 312000
P. R. China
e-mail: bhsheng@usx.edu.cn

YE PEIXIN, HAN YONGJIE
School of Mathematics and LPMC
Nankai University
Tianjin 300071
P. R. China
corresponding author: yepx@nankai.edu.cn

Abstract: In the present paper,we investigate the learning rates of regularized regression with sample dependent hypothesis and p-loss. We show the robustness of the solutions with respect to the probability distributions, with which we provide the sample error. Also,we show the approximation error with a kind of $K-$functional whose convergence rates are described in possibility. Finally, we show the explicit learning rates in cases of the norm regularization and the coefficient regularization. The results show that the parameters p have influences on the learning rates.

Key–Words: Regularized regressions,learning rates, reproducing kernel Hilbert spaces,sample dependent,p-loss.

1 Introduction

It is known that the performance of the regularized regression with least square loss and sample dependent hypothesis spaces has been studied fully (see, e.g., [5, 7, 12, 15, 18, 20, 22, 23, 24, 25, 27]). The aim of the present paper is to give an investigation on the learning rates of regularized regression with p-loss.

Let X be a compact metric space and $Y = \Re$. Let $\rho(x,y) = \rho(y|x)\rho_X(x)$ be a fixed but unknown Borel probability distribution on $Z := X \times Y$ with $\rho_X(x)$ being the marginal distribution and $\rho(y|x)$ the conditional probability distribution.

Let $V(t) : \Re \to \Re_+$ be a convex loss function and $\mathcal{E}_{\rho,V}(f) = \int_Z V(|y - f(x)|)\, d\rho$ be the integral error. The minimizer f_V^* defined by

$$f_V^*(x) := \arg\min_f \mathcal{E}_{\rho,V}(f)$$

over all measurable functions controls the relation between x and y. In particular, if $V(t) = t^2$ is the least square loss, then, f_V^* is exactly the regression function $f_\rho(x) = E(y|x) = \int_Y y\, d\rho(y|x)$(see [5]). The task of learning theory is to find, from the sample $z = (z_i)_{i=1}^m = ((x_i,y_i))_{i=1}^m \in Z^m$ drawn independent and identically according to the unknown probability distribution $\rho(x,y)$, a function f_z which is a good approximation of f_V^*.

A way of obtaining f_z is the following empirical Tikhonov regularization regressions (see e.g. [6, 8])

$$f_{z,\lambda,V,\mathcal{H}} = \arg\min_{f\in\mathcal{H}} (\frac{1}{m}\sum_{i=1}^m V(|y_i - f(x_i)|) + \lambda\, \Omega(f)), \quad (1)$$

where λ are positive constants called the regularized parameters, $\mathcal{H}$ is a hypothesis space which is, in most of the case, a function space consisting of functions on X with norm $\|\cdot\|_{\mathcal{H}}$, $\Omega(f) : \mathcal{H} \to R$ is the penalty function called regularizer.

When $\mathcal{H}$ are reproducing kernel Hilbert spaces(RKHS)(see [1, 2]), the learning rates have been studied by many papers (see e.g. [3, 4, 6, 8, 11, 18, 23]). In particular, for $V(t) = V_p(t) = |t|^p$, [3] shows that

$$\mathcal{E}_{\rho,V_p}(f_{z,\lambda,V_p,\mathcal{H}_K}) \to \mathcal{E}_{\rho,V_p}(f_{V_p}^*)(m \to +\infty) \quad (2)$$

if $\int_Z |y|^p\ d\rho < +\infty$, $\lambda^{p^*}m \to +\infty$ and $p^* = \max\{2p, p^2\}$.

It is necessary for us to show the convergence rates for (2).

Let $K(x,y) = K_x(y) : X \times X \to \Re$ be continuous, symmetric and positive semi-definite, i.e., for any finite set of distinct points $\overline{X} = \{x_1, x_2, \cdots, x_l\} \subset X$, the matrix $K_{\overline{X},\overline{X}} = (K(x_i,x_j))_{i,j=1}^l$ is positive semi-definite. Then we call $K(x,y)$ a Mercer kernel on X.

Let $\overline{X} = \{x_1,\ x_2,\ \cdots,\ x_m\} \subset X$ be taken from the sample set $z = \{(x_i,y_i)\}_{i=1}^m$. Then, we can define on $\mathcal{H}_{K,\overline{X}} = \{f_\alpha(x) = \sum_{j=1}^{m} \alpha_j \ K_{x_j}(x) : \alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)^\top \in \Re^m\}$ an inner product $\langle \cdot, \cdot \rangle_{K,\overline{X}}$ on $\mathcal{H}_{K,\overline{X}}$ satisfying $\langle K_t, K_s \rangle_{K,\overline{X}} = K(t,s)$ for $t, s \in \overline{X}$ which yields

$$\|f_\alpha\|^2_{K,\overline{X}} = \alpha^\top \ K_{\overline{X},\overline{X}} \ \alpha$$

and the reproducing property (see[20])

$$f(x) = \langle f, K_x \rangle_{K,\overline{X}}, \qquad x \in \overline{X}. \tag{3}$$

Take $\mathcal{H} = \mathcal{H}_{K,\overline{X}}$ in (1). Then, we have the following sample dependent framework with a general loss $V(t)$

$$\begin{aligned} f_{z,\lambda,V} &= f_{\alpha_{z,\lambda,V}}, \\ \alpha_{z,\lambda,V} &= \min_{\alpha \in R^m} (\frac{1}{m} \sum_{i=1}^{m} V(|y_i - f_\alpha(x_i)|) \\ &\quad + \lambda \ \Omega(f_\alpha)). \end{aligned} \tag{4}$$

The integral framework with respect to (4) is

$$\begin{aligned} f_{\rho,\lambda,V} &= f_{\alpha_{\rho,\lambda,V}}, \\ \alpha_{\rho,\lambda,V} &= \min_{\alpha \in R^m} (\mathcal{E}_{\rho,V}(f_\alpha) + \lambda \ \Omega(f_\alpha)). \end{aligned} \tag{5}$$

It has been pointed by [25] and [26] that, the error analysis for algorithm (4) is essentially different from and more difficult than that for (1) since the dependent of $\mathcal{H}_{K,\overline{X}}$ on the samples $\overline{X}$. For example, when we estimate the error with capacity approach, the sample dependent nature of the algorithm will lead an extra error term called hypothesis error which makes the analysis become more complexity. When $V(t)$ are Lipschitz loss,the convergence analysis for scheme (4) has been studied in ([10, 11, 14]). When $\Omega(f_\alpha) = \sum_{i=1}^{m} \alpha_i^2$ and $V(t) = t^2$, the learning rates are estimated in [21];When $\Omega(f_\alpha) = \alpha^\top \ K_{\overline{X},\overline{X}} \ \alpha$ and $V(t) = t^2$, the learning rates are estimated in [24]; When $\Omega(f_\alpha) = \sum_{i=1}^{m} |\alpha_i|$ and $V(t) = t^2$, the learning rates are shown in [17] and [26];[22] provides the learning rates for $V(t) = t^2$ and $\Omega(f_\alpha) = \sum_{i=1}^{m} |\alpha_i|^p (1 < p < 2)$. When $V(t) = |t|^p$ and $\Omega(f_\alpha) = \|f_\alpha\|^2_K$, we have

$$\begin{aligned} f_{z,\lambda,p} &= f_{\alpha_{z,\lambda,p}}, \\ \alpha_{z,\lambda,p} &= arg \min_{\alpha \in \Re^m} (\frac{1}{m} \sum_{i=1}^{m} |y_i - f_\alpha(x_i)|^p \\ &\quad + \lambda \ \alpha^\top \ K_{\overline{X},\overline{X}} \ \alpha). \end{aligned} \tag{6}$$

When $V(t) = |t|^p$ and $\Omega(f_\alpha) = m \sum_{i=1}^{m} |\alpha_i|^2$,we have

$$\begin{aligned} f^*_{z,\lambda,p} &= f_{\alpha^*_{z,\lambda,p}}, \\ \alpha^*_{z,\lambda,p} &= arg \min_{\alpha \in \Re^m} (\frac{1}{m} \sum_{i=1}^{m} |y_i - f_\alpha(x_i)|^p \\ &\quad + \lambda \ m \sum_{i=1}^{m} |\alpha_i|^2). \end{aligned} \tag{7}$$

Since $V(t) = |t|^p$ is local Lipschitz loss,the method used in [10] and [13] can not be used. The integral operator approach will not be used as well since the explicit solutions of (6) and (7) can not be obtained,we need to show the performance in another way.

In the present paper, we define the general integral regularized framework with respect to scheme (6) and (7) respectively by

$$\begin{aligned} &\alpha^{(\rho)}_p := \alpha^{(\rho)}_{\lambda,p} \\ &= \arg \min_{\alpha \in \Re^m} (\mathcal{E}^{(p)}_\rho(f_\alpha) + \lambda \ \alpha^\top \ K_{\overline{X},\overline{X}} \ \alpha), \end{aligned} \tag{8}$$

and

$$\begin{aligned} &\alpha^{(\rho),*}_p := \alpha^{(\rho),*}_{\lambda,p} \\ &= \arg \min_{\alpha \in \Re^m} (\mathcal{E}^{(p)}_\rho(f_\alpha) + \lambda \ m \sum_{i=1}^{m} |\alpha_i|^2), \end{aligned} \tag{9}$$

where $\mathcal{E}^{(p)}_\rho(f) = \int_Z |y - f(x)|^p d\rho$. Take

$$f^{(p)}_\rho = \arg \min_f \mathcal{E}^{(p)}_\rho(f), \tag{10}$$

where the minimize is taken over all the measurable functions. Define the empirical measure $\rho_z(x,y)$ corresponding to the sample z and the bounded $\rho-$measurable function $f(x,y)$ on Z by

$$\int_Z f(x,y) \ d\rho_z = \frac{1}{m} \sum_{i=1}^{m} f(x_i, y_i). \tag{11}$$

Let $L^p(\rho_X)$ denote the class of ρ_X-integrable functions $f(x)$ which satisfy

$$\|f\|_{L^p(\rho_X)} = (\int_X |f(x)|^p \ d\rho_X)^{\frac{1}{p}} < +\infty.$$

Then, we know $\| \cdot \|_{L^p(\rho_X)}$ satisfies, for any $f, g \in L^p(\rho_X)$,

$$|\|f\|_{L^p(\rho_X)} - \|g\|_{L^p(\rho_X)}| \leq \|f - g\|_{L^p(\rho_X)}.$$

It follows for $f, g \in L^p(\rho_X)$ that, for any $f, g \in L^p(\rho_X)$,

$$|\mathcal{E}_\rho^{(p)}(f)^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(g)^{\frac{1}{p}}| \leq \|f-g\|_{L^p(\rho_X)}. \quad (12)$$

Let $\Re^m$ be the $m-$dimensional Euclidean space, $f(x) : \Re^m \to \Re$ be a differentiable convex function and $\nabla f(x) = (\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_m})$ be the usual gradient. Then,

$$f(x') - f(x) \geq \langle \nabla f(x), x' - x \rangle, \ x, x' \in R^m. \quad (13)$$

A well known result is, if $f(x)$ is a convex function on X, then x_0 is the minimal value point of $f(x)$ on X if and only if $\nabla f(x_0) = 0$.

Since X is a compact set,we have by the reproducing property (3) that for any $f \in \mathcal{H}_{K,\overline{X}}$

$$\begin{aligned} |f(x)| &\leq \|f\|_{K,\overline{X}} \times \|K_x\|_{K,\overline{X}} \\ &= \|f\|_{K,\overline{X}} \times \sqrt{K(x,x)}. \end{aligned}$$

It follows by $\int_X d\rho_X = 1$ that

$$\|f\|_{L^p(\rho_X)} = (\int_X |f(x)|^p d\rho_X)^{\frac{1}{p}} \leq k\|f\|_{K,\overline{X}}. \quad (14)$$

Then, by the definitions of $\alpha_p^{(\rho)}$,(12) and (14) we have

$$\begin{aligned} &\mathcal{E}_\rho^{(p)}(f_{\alpha_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & |\mathcal{E}_\rho^{(p)}(f_{\alpha_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}})^{\frac{1}{p}}| \\ &+\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & \|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{L^p(\rho_X)} \\ &+\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & k\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}} \\ &+\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}}. \quad (15) \end{aligned}$$

On the other hand, define on $\Re^m (m \geq 1)$ the norm

$$\|a\|^2 = \sum_{i=1}^m |a_i|^2 = a^\top, a = (a_1, a_2, \cdots, a_m)^\top \in \Re^m,$$

and for $a = (a_1, a_2, \cdots, a_m)^\top \in \Re^m$ and $b = (b_1,\ b_2,\ \cdots,\ b_m)^\top \in \Re^m$ define the inner product

$$\langle a, b \rangle = \sum_{i=1}^m a_i\ b_i = a^\top\ b.$$

Then,

$$\begin{aligned} \max_{x\in X} |f_\alpha(x)| &= \max_{x\in X} |\sum_{k=1}^m \alpha_k K(x_k, x)| \\ &\leq k\sqrt{m}\|\alpha\|. \quad (16) \end{aligned}$$

It follows by the definitions of $\alpha_p^{(\rho),*}$, (12) and (16) that

$$\begin{aligned} &\mathcal{E}_\rho^{(p)}(f_{\alpha^*_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & |\mathcal{E}_\rho^{(p)}(f_{\alpha^*_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}})^{\frac{1}{p}}| \\ &+\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & k\sqrt{m}\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\| \\ &+\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}}. \quad (17) \end{aligned}$$

Then, to bound the learning rates of the algorithms (6) and (7), we need to provide the convergence rates for the sample errors $\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}}$ and $\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\|$ and,bound the convergence rates for the approximation errors $\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}}$ and $\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}}$.Along this line,we bound the learning errors for algorithm (6) and algorithm (7).

The learning rates of algorithm (6) are estimated as following Theorem 1.

Theorem 1 *Let* $p > 1$, $\alpha_{z,\lambda,p}$ *be defined as in (6), Mercer kernel* $K(x,y)$ *satisfy*

$$\sup_{(x,y)\in X\times X} |K(x,y)| < +\infty,$$

$\rho(x,y)$ *satisfy*

$$|\rho|_p = \int_X |y|^p\ d\rho < +\infty.$$

Then, there is a constant c_p *depending only on* p *such that for any* $0 < \delta < 1$, *with confidence* $1 - 2\delta$, *there holds*

$$\begin{aligned} &\mathcal{E}_\rho^{(p)}(f_{\alpha_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}} \frac{kp}{\lambda m^{q^*}} + \frac{E(K(f_\rho^{(p)}, \lambda))}{\delta}, (18) \end{aligned}$$

where $k = \sup_{(x,y)\in X\times X} |K(x,y)|$, $q^* = \min(\frac{1}{2}, \frac{1}{p})$ *and*

$$\begin{aligned} K(f_\rho^{(p)}, \lambda) &= \Big[\inf_{\alpha\in R^m} (\mathcal{E}_\rho^{(p)}(f_\alpha) + \lambda\|f_\alpha\|^2_{K,\overline{X}})\Big]^{\frac{1}{p}} \\ &\quad -\mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \end{aligned}$$

and

$$E(f) = \int_{X^m} f(x_1, x_2, \cdots, x_m) d\rho_X(x_1) \cdots d\rho_X(x_m).$$

Also, the learning rates of algorithm (7) are estimated as following Theorem 2.

Theorem 2 *Let $p > 1, \alpha^*_{z,\lambda,p}$ be defined as in (7), Mercer kernel $K(x,y)$ satisfy $\sup\limits_{(x,y)\in X\times X} |K(x,y)| < +\infty$, $\rho(x,y)$ satisfy $|\rho|_p = \int_X |y|^p d\rho < +\infty$. Then, there is a constant c_p depending only on p such that for any $0 < \delta < 1$, with confidence $1-2\delta$, there holds*

$$\begin{aligned}&\mathcal{E}^{(p)}_\rho(f_{\alpha^*_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}^{(p)}_\rho(f^{(p)}_\rho)^{\frac{1}{p}}\\ \le\ & \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}}\times\frac{kp}{\lambda m^{q^*}}+\frac{E(K^*(f^{(p)}_\rho,\lambda))}{\delta}, \quad (19)\end{aligned}$$

where

$$\begin{aligned}K^*(f^{(p)}_\rho,\lambda) &= \Big[\inf_{\alpha\in R^m}(\mathcal{E}^{(p)}_\rho(f_\alpha)+m\lambda\|\alpha\|^2)\Big]^{\frac{1}{p}}\\ &\quad -\mathcal{E}^{(p)}_\rho(f^{(p)}_\rho)^{\frac{1}{p}}\end{aligned}$$

Theorem 1 and Theorem 2 will be proved in Section 3.

It seems that the estimates (18) and (19) are neither explicit nor completeness since the parts $\frac{E(K(f^{(p)}_\rho,\lambda))}{\delta}$ and $\frac{E(K^*(f^{(p)}_\rho,\lambda))}{\delta}$. However,if $f^{(p)}_\rho$ is in the range of integral operator $L_K(f)$ defined by

$$L_K(f,x) = \int_X f(u)K(u,x)d\rho_X(u), \quad x\in X,$$

then,we have the following explicit estimates for (18).

Theorem 3 *Let $p > 1, \alpha_{z,\lambda,p}$ be defined as in (6), Mercer kernel $K(x,y)$ satisfy $\sup\limits_{(x,y)\in X\times X} |K(x,y)| < +\infty$, $\rho(x,y)$ satisfy $|\rho|_p = \int_X |y|^p\ d\rho < +\infty$. If there is a function $\varphi \in L^2(\rho_X)$ such that $f^{(p)}_\rho(x) = L_K(\varphi,x)$, then, there is a constant c_p depending only on p such that for any $0 < \delta < 1$, with confidence $1-2\delta$, there holds*

$$\begin{aligned}&\mathcal{E}^{(p)}_\rho(f_{\alpha_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}^{(p)}_\rho(f^{(p)}_\rho)^{\frac{1}{p}}\\ &\le \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}}\times\frac{kp}{\lambda m^{q^*}}\\ &+\frac{1}{\delta}\Big(k\sqrt{\frac{\nu-\mu}{m}}+\sqrt[p]{\lambda\nu}\Big), \quad (20)\end{aligned}$$

where

$$\begin{aligned}\nu &= \int_X \varphi(y)^2K(y,y)d\rho_X(y),\\ \mu &= \int_X\int_X \varphi(x)\varphi(y)K(x,y)\ d\rho_X(x)d\rho_X(y).\end{aligned}$$

Also, we provide an explicit estimate for (19) by the following Theorem 4.

Theorem 4 *Let $p > 1, \alpha^*_{z,\lambda,p}$ be defined as in (7), Mercer kernel $K(x,y)$ satisfy $\sup\limits_{(x,y)\in X\times X} |K(x,y)| < +\infty$, $\rho(x,y)$ satisfy $|\rho|_p = \int_X |y|^p\ d\rho < +\infty$. If there is a function $\varphi\in L^2(\rho_X)$ such that $f^{(p)}_\rho(x) = L_K(\varphi,x)$, then, there is a constant c_p depending only on p such that for any $0 < \delta < 1$, with confidence $1-2\delta$, there holds*

$$\begin{aligned}&\mathcal{E}^{(p)}_\rho(f_{\alpha^*_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}^{(p)}_\rho(f^{(p)}_\rho)^{\frac{1}{p}}\\ &\le \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}}\times\frac{kp}{\lambda m^{q^*}}\\ &+\frac{1}{\delta}(k\sqrt{\frac{\nu-\mu}{m}}+\sqrt[p]{\lambda\|\varphi\|_{L^2(\rho_X)}}). \quad (21)\end{aligned}$$

2 Some lemmas

To show Theorem 1-Theorem 4,we need some lemmas.

Lemma 5 *Let $\alpha^{(\rho)}_p$ be the unique solution of scheme (8). Then, for any $f\in\mathcal{H}_{K,\overline{X}}$, holds*

$$\begin{aligned}&2\lambda\langle f, f_{\alpha^{(\rho)}_p}\rangle_{K,\overline{X}}\\ =\ & p\langle f, \int_Z (y-f_{\alpha^{(\rho)}_p}(x))^{p-1}\\ &\times sgn(y-f_{\alpha^{(\rho)}_p}(x))\ K_x(\cdot)\ d\rho\rangle_{K,\overline{X}} \quad (22)\end{aligned}$$

and

$$\mathcal{E}^{(p)}_\rho(f_{\alpha^{(\rho)}_p}) \le |\rho|_p. \quad (23)$$

If $\alpha^{(\rho),*}_p$ is the unique solution of scheme (9), then,

$$\begin{aligned}&2\lambda\ m\ \alpha^{(\rho),*}_p\\ =\ & p\int_Z |y-f_{\alpha^{(\rho),*}_p}(x)|^{p-1}\\ &\times sgn(y-f_{\alpha^{(\rho),*}_p}(x))\ K_{\overline{X}}(x)\ d\rho \quad (24)\end{aligned}$$

and

$$\mathcal{E}^{(p)}_\rho(f_{\alpha^{(\rho),*}_p}) \le |\rho|_p, \quad (25)$$

where and in the whole paper we write

$$K_{\overline{X}}(x) = (K(x_1,x),K(x_2,x),\cdots,K(x_m,x))$$

and for a vector function $f(x) = (f_1(x),\ \cdots,\ f_m(x))^\top$ and a real function $\alpha(x)$ on X we define

$$f(x)\alpha(x) = (f_1(x)\alpha(x),\cdots,f_m(x)\alpha(x))^\top$$

and

$$\int_X f(x)\,\alpha(x)\,d\rho_X = \Big(\int_X f_1(x)\,\alpha(x)\,d\rho_X,\cdots,\int_X f_m(x)\,\alpha(x)\,d\rho_X\Big)^\top.$$

Proof. Since $\lambda>0$, $\|f_\alpha\|^2_{K,\overline{X}}$ and $\|\alpha\|^2$ are strict convex functions about α on $\Re^m$, $V_p(t)=|t|^P$ is a strict convex function on R, we know (8) and (9) are strict convex optimization problems about α on $\Re^m$. Both $\alpha_p^{(\rho)}$ and $\alpha_p^{(\rho),*}$ are uniqueness.

We now show (22). By the Theorem 4.2.1 in [9] we have the following results:

Let $A:\Re^q\to\Re^q$ be an affine mapping ($Ax=A_0x+b$) with A_0 linear and $b\in\Re^q$ and let g be a finite convex function on $\Re^q$. Then,

$$\nabla_x(g\circ A)(x)=A_0^*\,\nabla_{Ax}g(Ax) \tag{26}$$

for all $x\in\Re^q$, where A_0^* is the adjoint of A_0.

Since $|y-f_\alpha(x)|^p=|y-K_{\overline{X}}(x)\alpha|^p$, by taking $A_0=K_{\overline{X}}(x)$ and $b=y$ in (26), we have

$$\begin{aligned}\nabla_\alpha|y-f_\alpha(x)|^p &= -p|y-K_{\overline{X}}(x)\alpha|^{p-1}\\ &\quad\times sgn(y-f_\alpha(x))\,K_{\overline{X}}(x).\end{aligned}$$

Since $\alpha_p^{(\rho)}$ is the minimizer of (8) and

$$\nabla_\alpha(\alpha^\top K_{\overline{X},\overline{X}}\,\alpha)|_{\alpha=\alpha_p^{(\rho)}}=2\,K_{\overline{X},\overline{X}}\,\alpha_p^{(\rho)},$$

we have

$$\begin{aligned}0 &= \nabla_\alpha\Big(\int_Z|y-f_\alpha(x)|^p\,d\rho\Big)|_{\alpha=\alpha_p^{(\rho)}}\\ &\quad+2\lambda\;K_{\overline{X},\overline{X}}\,\alpha_p^{(\rho)}\\ &= \int_Z\nabla_\alpha|y-f_\alpha(x)|^p|_{\alpha=\alpha_p^{(\rho)}}\,d\rho+2\lambda\,K_{\overline{X},\overline{X}}\,\alpha_p^{(\rho)}\\ &= -p\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\\ &\quad\times K_{\overline{X}}(x)\,d\rho+2\lambda\,K_{\overline{X},\overline{X}}\,\alpha_p^{(\rho)}.\end{aligned}$$

Hence,

$$\begin{aligned}&2\lambda\,K_{\overline{X},\overline{X}}\,\alpha_p^{(\rho)}\\ &=p\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\\ &\times K_{\overline{X}}(x)\,d\rho.\end{aligned} \tag{27}$$

By the definition of the matrix $K_{\overline{X},\overline{X}}$ we have

$$\begin{aligned}&2\lambda(f_{\alpha_p^{(\rho)}}(x_1),\cdots,f_{\alpha_p^{(\rho)}}(x_m))\\ &= p\Big(\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\\ &\quad\times K_x(x_1)\,d\rho,\cdots,\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\\ &\quad\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\,K_x(x_m)\,d\rho\Big).\end{aligned} \tag{28}$$

By the reproducing property (3) we have $f_{\alpha_p^{(\rho)}}(x_i)=\langle f_{\alpha_p^{(\rho)}},K_{x_i}\rangle_{K,\overline{X}}$ and

$$K_x(x_i)=\langle K_x,K_{x_i}\rangle_K,\qquad i=1,2,\cdots,m.$$

Then, (28) yields

$$\begin{aligned}&2\lambda\Big(\langle f_{\alpha_p^{(\rho)}},K_{x_1}\rangle_{K,\overline{X}},\cdots,\langle f_{\alpha_p^{(\rho)}},K_{x_m}\rangle_{K,\overline{X}}\Big)\\ &= p\Big(\Big\langle\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\\ &\quad\times K_x(\cdot)\,d\rho,K_{x_1}\Big\rangle_K,\cdots,\Big\langle\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\\ &\quad\times sgn(y-f_{\alpha_p^{(\rho)}}(x))K_x(\cdot)\,d\rho,K_{x_m}\Big\rangle_{K,\overline{X}}\Big).\end{aligned}$$

Therefore, we have for $i=1,2,\cdots,m$, that

$$\begin{aligned}&2\lambda\langle f_{\alpha_p^{(\rho)}},K_{x_i}\rangle_{K,\overline{X}}\\ &= p\Big\langle\int_Z|y-f_{\alpha_p^{(\rho)}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho)}}(x))\\ &\quad\times K_x(\cdot)\,d\rho,K_{x_i}\Big\rangle_{K,\overline{X}}.\end{aligned}$$

(22) then holds.

We now show (24). Since $\alpha_p^{(\rho),*}$ is the unique solution of (9),we have

$$\begin{aligned}0 &= \nabla_\alpha\Big(\int_Z|y-f_\alpha(x)|^p\,d\rho\Big)|_{\alpha=\alpha_p^{(\rho),*}}\\ &\quad+2\lambda m\,\|\alpha\|^2)|_{\alpha=\alpha_p^{(\rho),*}}\\ &= -p\int_Z|y-f_{\alpha_p^{(\rho),*}}(x)|^{p-1}\times sgn(y-f_{\alpha_p^{(\rho),*}}(x))\\ &\quad\times K_{\overline{X}}(x)\,d\rho+2\lambda m\alpha_p^{(\rho),*}.\end{aligned}$$

Therefore, (24) holds.

Take $\alpha=0$ in (8),we have (23).Take $\alpha=0$ in (9),we have (25). □

Lemma 6 *For any $f,g\in\mathcal{H}_{K,\overline{X}}$ there holds*

$$\begin{aligned}&\|f\|^2_{K,\overline{X}}-\|g\|^2_{K,\overline{X}}\\ &= 2\langle f-g,\;g\rangle_{K,\overline{X}}+\|f-g\|^2_{K,\overline{X}}\end{aligned} \tag{29}$$

and for any $a,b\in R^m$ there holds

$$\|a\|^2-\|b\|^2=2\langle a-b,\;b\rangle+\|a-b\|^2. \tag{30}$$

Proof. (29) and (30) may be obtained by the parallelogram identity. □

We now show the following robustness for schemes (8) and (9).

Lemma 7 *Let ρ and γ be two given Borel probability distributions on Z. Then, We have the following robustness for the solutions of (8) and (9):*

(1) if $\alpha_p^{(\rho)}$ and $\alpha_p^{(\gamma)}$ are the solutions of (8) with respect to ρ and γ respectively,then,

$$\begin{aligned} & \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|_{K,\overline{X}} \\ \leq\ & \frac{p}{\lambda}\Big\| \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\ & \times K_x(\cdot)\, d\rho \\ & - \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\ & \times K_x(\cdot)\, d\gamma \Big\|_{K,\overline{X}}. \end{aligned} \tag{31}$$

(2) if $\alpha_p^{(\rho),*}$ and $\alpha_p^{(\gamma),*}$ are the solutions of (9) with respect to ρ and γ respectively,then,

$$\begin{aligned} & \|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\| \\ \leq\ & \frac{p}{\lambda m}\Big\| \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\ & \times K_{\overline{X}}(x)\, d\rho \\ & - \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\ & \times K_{\overline{X}}(x)\, d\gamma \Big\|. \end{aligned} \tag{32}$$

Proof. *Proof of formula (31).* Since $(|t|^p)' = p|t|^{p-1} sgn(t)$,we have by the convexity of the function $|t|^p$ that

$$|x|^p - |y|^p \geq p|y|^{p-1} sgn(y)\ (x-y),\ x, y \in \Re^1. \tag{33}$$

Then, the reproducing property gives

$$f_{\alpha_p^{(\gamma)}}(x) - f_{\alpha_p^{(\rho)}}(x) = \Big\langle K_x, f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}} \Big\rangle_{K,\overline{X}}, x \in X$$

and

$$\begin{aligned} & |y - f_{\alpha_p^{(\gamma)}}(x)|^p - |y - f_{\alpha_p^{(\rho)}}(x)|^p \\ \geq\ & p|y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\ & \times (f_{\alpha_p^{(\rho)}}(x) - f_{\alpha_p^{(\gamma)}}(x)) \\ =\ & \Big\langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ -p|y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\ & \times sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x \Big\rangle_{K,\overline{X}}. \end{aligned} \tag{34}$$

and hence (34) gives

$$\begin{aligned} & \int_Z |y - f_{\alpha_p^{(\gamma)}}(x)|^p d\gamma - \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^p\, d\gamma \\ \geq\ & \Big\langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ -p \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\ & \times sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x(\cdot)\ d\gamma \Big\rangle_{K,\overline{X}}. \end{aligned}$$

Take $f = f_{\alpha_p^{(\rho)}}$ and $g = f_{\alpha_p^{(\gamma)}}$ into (29). Then,

$$\begin{aligned} & \|f_{\alpha_p^{(\rho)}}\|^2_{K,\overline{X}} - \|f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}} \\ =\ & 2\langle f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}},\ f_{\alpha_p^{(\gamma)}} \rangle_{K,\overline{X}} \\ & + \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}}. \end{aligned} \tag{35}$$

(35) gives

$$\begin{aligned} & (\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\gamma)}}) + \lambda\, \|f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}}) - (\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\rho)}}) \\ & + \lambda\, \|f_{\alpha_p^{(\rho)}}\|^2_{K,\overline{X}}) \\ \geq\ & \langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ -p \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\ & \times sgn(y - f_{\alpha_p^{(\rho)}}(x)) K_x(\cdot)\ d\gamma \rangle_{K,\overline{X}} \\ & + 2\lambda\, \langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ f_{\alpha_p^{(\rho)}} \rangle_{K,\overline{X}} \\ & + \lambda\, \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}} \\ =\ & \Big\langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ 2\lambda\, f_{\alpha_p^{(\rho)}} - p \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\ & \times sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x(\cdot)\ d\gamma \Big\rangle_{K,\overline{X}} \\ & + \lambda\, \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}} \\ =\ & p\langle f_{\alpha_p^{(\gamma)}} - f_{\alpha_p^{(\rho)}},\ \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\ & \times sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x(\cdot)\ d\rho \\ & - \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1}\ sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\ & \times K_x(\cdot)\ d\gamma \rangle_{K,\overline{X}} \\ & + \lambda\ \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}}, \end{aligned} \tag{36}$$

where, in the last deduction, we have used (22). By the definitions of $\alpha_p^{(\rho)}$ and $\alpha_p^{(\gamma)}$ we have

$$\begin{aligned} & (\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\gamma)}}) + \lambda \|f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}}) - (\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\rho)}}) \\ & + \lambda\, \|f_{\alpha_p^{(\rho)}}\|^2_{K,\overline{X}}) \leq 0, \end{aligned}$$

which and (36) give

$$\lambda\, \|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|^2_{K,\overline{X}}$$

$$\begin{aligned}
&\le p\langle f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}, \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\
&\quad \times sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x(\cdot)\ d\rho \\
&\quad - \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\
&\quad \times K_x(\cdot)\ d\gamma\rangle_{K,\overline{X}} \\
&\le p\|f_{\alpha_p^{(\rho)}} - f_{\alpha_p^{(\gamma)}}\|_{K,\overline{X}} \\
&\quad \times \| \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\
&\quad \times K_x(\cdot)\ d\rho \\
&\quad - \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\
&\quad \times K_x(\cdot)\ d\gamma\|_{K,\overline{X}}.
\end{aligned}$$

(31) thus holds.

Proof of the formula (32). The equality

$$f_{\alpha_p^{(\gamma),*}}(x) - f_{\alpha_p^{(\rho),*}}(x) = \langle K_{\overline{X}}(x),\ \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}\rangle$$

and the inequality (33) yield

$$\begin{aligned}
&|y - f_{\alpha_p^{(\gamma),*}}(x)|^p - |y - f_{\alpha_p^{(\rho),*}}(x)|^p \\
\ge\ & p|y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\
& \times (f_{\alpha_p^{(\rho),*}}(x) - f_{\alpha_p^{(\gamma),*}}(x)) \\
=\ & \langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*},\ -p|y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) K_{\overline{X}}(x)\rangle. \qquad (37)
\end{aligned}$$

From (37) we get that

$$\begin{aligned}
&\int_Z |y - f_{\alpha_p^{(\gamma),*}}(x)|^p d\gamma - \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^p d\gamma \\
\ge\ & \langle \alpha_p^{(\gamma)} - \alpha_p^{(\rho),*},\ -p\int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) K_{\overline{X}}(x)\ d\gamma\rangle.
\end{aligned}$$

Taking $a = \alpha_p^{(\gamma),*}$ and $b = \alpha_p^{(\rho),*}$ into (30), we have

$$\begin{aligned}
&\|\alpha_p^{(\gamma),*}\|^2 - \|\alpha_p^{(\rho),*}\|^2 \\
=\ & 2\langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}, \alpha_p^{(\rho),*}\rangle + \|\alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}\|^2.
\end{aligned}$$

Above two equalities give

$$\begin{aligned}
&(\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\gamma),*}}) + \lambda m\ \|\alpha_p^{(\gamma),*}\|^2) \\
&-(\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\rho),*}}) + \lambda m\ \|\alpha_p^{(\rho),*}\|^2) \\
\ge\ & \langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}, -p\int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) K_{\overline{X}}(x) d\gamma\rangle \\
& + 2\lambda m\ \langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}, \alpha_p^{(\rho),*}\rangle \\
& + \lambda m\|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\|^2 \\
=\ & \langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}, 2\lambda m \alpha_p^{(\rho),*} \\
& -p\int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\
& \times K_{\overline{X}}(x) d\gamma\rangle \\
& + \lambda m\ \|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\|^2 \\
=\ & p\langle \alpha_p^{(\gamma),*} - \alpha_p^{(\rho),*}, \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) K_{\overline{X}}(x) d\rho \\
& - \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1}\ sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\
& \times K_{\overline{X}}(x)\ d\gamma\rangle \\
+\ & \lambda\ m\ \|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\|^2, \qquad (38)
\end{aligned}$$

where, in the last deduction, we have used (24). By the definitions of $\alpha_p^{(\rho),*}$ and $\alpha_p^{(\gamma),*}$ we have

$$\begin{aligned}
&(\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\gamma),*}}) + \lambda m\|\alpha_p^{(\gamma),*}\|^2) - (\mathcal{E}_\gamma^{(p)}(f_{\alpha_p^{(\rho),*}}) \\
&+ \lambda m\ \|\alpha_p^{(\rho),*}\|^2) \le 0,
\end{aligned}$$

which and (36) give

$$\begin{aligned}
&\lambda m\ \|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\|^2 \\
\le\ & p\langle \alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*},\ \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x))\ K_{\overline{X}}(x)\ d\rho \\
& - \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x))\ K_{\overline{X}}(x)\ d\gamma\rangle \\
\le\ & p\|\alpha_p^{(\rho),*} - \alpha_p^{(\gamma),*}\| \times \| \int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} \\
& \times sgn(y - f_{\alpha_p^{(\rho),*}}(x))\ K_{\overline{X}}(x)\ d\rho \\
& - \int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\
& \times K_{\overline{X}}(x)\ d\gamma\|.
\end{aligned}$$

(32) then holds.

We now give an estimate for the sample error $\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}}$.

Lemma 8 *Let* $p > 1, \alpha_p^{(\rho)}$ *and* $\alpha_{z,\lambda,p}$ *be defined as in (8) and (6) respectively. Then,there is a positive constant* c_p *such that, for any* $0 < \delta < 1$, *with confidence* $1 - \delta$, *holds*

$$\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}} \le \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}} \times \frac{kp}{\lambda m^{q^*}}. \qquad (39)$$

Proof. Notice the following large number law (see [3]): *Let Z be a measurable space,ρ be a distribution on Z, H be a Hilbert space and $f : Z \to H$ be a measurable function with $\|f\|_q = (E(\|f\|_H^q))^{\frac{1}{q}} < +\infty$ for $q \in (1, +\infty)$. We write $q^* := min(1/2, (q-1)/q)$.Then,there exists a universal constant $c_q > 0$ such that for all $\varepsilon > 0$ and all $m \geq 1$ we have*

$$Prob^m((z_1, z_2, \cdots, z_m) \in Z^m : \|\frac{1}{m}\sum_{i=1}^m f(z_i) - E(f)\|_H \geq \varepsilon) \leq c_q\Big(\frac{\|f\|_q}{\varepsilon m^{q^*}}\Big)^q$$

It follows that

$$P^m((z_1, z_2, \cdots, z_m) \in Z^m : \|\frac{1}{m}\sum_{i=1}^m f(z_i) - E(f)\|_H \leq \varepsilon) \geq 1 - c_q\Big(\frac{\|f\|_q}{\varepsilon m^{q^*}}\Big)^q$$

and for any $0 < \delta < 1$,with confidence $1 - \delta$, holds

$$\|\frac{1}{m}\sum_{i=1}^m f(z_i) - E(f)\|_H \leq \Big(\frac{c_q}{\delta}\Big)^{\frac{1}{q}}\frac{\|f\|_q}{m^{q^*}}. \quad (40)$$

Denoted by

$$\begin{aligned} & f(z) = f(x, y) \\ = \ & |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1} sgn(y - f_{\alpha_p^{(\rho)}}(x))\ K_x(\cdot). \end{aligned}$$

Then,

$$\begin{aligned} \|f\|_K &= |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1}\sqrt{K(x,x)} \\ &\leq k|y - f_{\alpha_p^{(\rho)}}(x)|^{p-1}. \end{aligned} \quad (41)$$

Take $q = \frac{p}{p-1}$. Then, by (23), one has

$$\begin{aligned} &\|f\|_q = (\int_Z \|f\|_K^q d\rho)^{\frac{1}{q}} \\ &\leq k(\int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^p d\rho_X)^{\frac{1}{q}} \leq k|\rho|_p^{\frac{1}{q}}. \end{aligned} \quad (42)$$

It follows by (40) and (42) that there is a positive constant c_p depending only upon p such that, with confidence $1 - \delta$, there holds

$$\begin{aligned} &\|\int_Z |y - f_{\alpha_p^{(\rho)}}(x)|^{p-1}\ sgn(y - f_{\alpha_p^{(\rho)}}(x)) \\ &\times K_x(\cdot)\ d\rho \\ &-\frac{1}{m}\sum_{i=1}^m |y_i - f_{\alpha_p^{(\rho)}}(x_i)|^{p-1}\ sgn(y_i - f_{\alpha_p^{(\rho)}}(x_i)) \\ &\times K_{x_i}(\cdot)\|_K \\ \leq\ & \sup_{f\in\mathcal{H}_K, \|f\|_q \leq k|\rho|_p^{\frac{1}{q}}} \|\frac{1}{m}\sum_{i=1}^m f(x_i, y_i) \\ &-\int_Z f(x,y)d\rho\|_K \\ \leq\ & \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}}\frac{k}{m^{q^*}}. \end{aligned} \quad (43)$$

Thus (43) together with (31) yield (39). Also, the sample error $\|\alpha_p^{(\rho),*} - \alpha_{z,\lambda,p}^*\|$ is bounded as the following Lemma 9.

Lemma 9 *Let $p > 1, \alpha_p^{(\rho),*}$ and $\alpha_{z,\lambda,p}^*$ be defined as in (9) and (7) respectively. Then, there is a positive constant c_p such that, for any $0 < \delta < 1$, with confidence $1 - \delta$, there holds*

$$\|\alpha_p^{(\rho),*} - \alpha_{z,\lambda,p}^*\| \leq \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}} \times \frac{kp}{\lambda m^{q^*+\frac{1}{2}}}. \quad (44)$$

Proof. Take

$$\begin{aligned} \xi(z) &= \xi(x,y) \\ &= |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1} \times sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\ &\quad \times K_{\overline{X}}(x). \end{aligned} \quad (45)$$

Then, $\xi(z) : Z \to R^m$ and

$$\begin{aligned} \|\xi(z)\| &= |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1}\sqrt{\sum_{k=1}^m K(x_k, x)^2} \\ &\leq k\sqrt{m}|y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1}. \end{aligned} \quad (46)$$

It follows that

$$\begin{aligned} \|\xi\|_q &= (\int_Z \|\xi(z)\|^q d\rho)^{\frac{1}{q}} \\ &= (\int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^p d\rho)^{\frac{1}{q}} \\ &\leq k\sqrt{m}|\rho|_p^{\frac{1}{q}}. \end{aligned} \quad (47)$$

We know by (40) and (47) that,there is a positive constant c_p depending only upon p such that, with confidence $1 - \delta$, holds

$$\begin{aligned} &\|\int_Z |y - f_{\alpha_p^{(\rho),*}}(x)|^{p-1}\ sgn(y - f_{\alpha_p^{(\rho),*}}(x)) \\ &\times K_x(\cdot)\ d\rho \\ &-\frac{1}{m}\sum_{i=1}^m |y_i - f_{\alpha_p^{(\rho),*}}(x_i)|^{p-1} \\ &\times sgn(y_i - f_{\alpha_p^{(\rho),*}}(x_i))\ K_{x_i}(\cdot)\| \\ \leq\ & \sup_{\|f\|_q \leq k\sqrt{m}|\rho|_p^{\frac{1}{q}}} \|\frac{1}{m}\sum_{i=1}^m f(x_i, y_i) - \int_Z f(x,y)d\rho\| \\ \leq\ & \Big(\frac{c_p|\rho|_p}{\delta}\Big)^{\frac{p-1}{p}}\frac{k}{m^{q^*-\frac{1}{2}}}. \end{aligned} \quad (48)$$

Inequality (48) together with (32) yield (44).

3 Proof of the Results

Proof of (18). By (14),(12) and the definition of $f_\rho^{(p)}$ we have

$$\begin{aligned} & \mathcal{E}_\rho^{(p)}(f_{\alpha_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & k\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}} + \Big(\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho)}}) \\ & +\lambda\|f_{\alpha_p^{(\rho)}}\|^2_{K,\overline{X}}\Big)^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ =\ & k\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}} + \Big(\inf_{\alpha\in R^m}(\mathcal{E}_\rho^{(p)}(f_\alpha) \\ & +\lambda\|f_\alpha\|^2_{K,\overline{X}})\Big)^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ =\ & k\|f_{\alpha_{z,\lambda,p}} - f_{\alpha_p^{(\rho)}}\|_{K,\overline{X}} + K(f_\rho^{(p)},\lambda) \\ =\ & A + B. \end{aligned} \tag{49}$$

On the other hand, we have the following Proposition 10.

Proposition 10 *([16]) Let $(\Omega,\mathcal{F},P)$ be a probability space,$z=(z_i)_{i=1}^m\in\Omega^m$ be samples drawn independently and identically (i.i.d.) according to ρ_X. $A(z)$ and $B(z)$ are functions about z on Ω^m. If there are $\varepsilon_1>0,\varepsilon_2>0$ such that for $0<\delta<1$,with confidence $1-\delta$,holds*

$$A(z)\leq\varepsilon_1, \qquad B(z)\leq\varepsilon_2,$$

then, with confidence $1-2\delta$,holds

$$A(z)+B(z)\leq\varepsilon_1+\varepsilon_2. \tag{50}$$

We now estimate B. By the Markov inequality we have

$$\begin{aligned} & Prob^m((z_1,z_2,\cdots,z_m)\in Z^m : B\geq\varepsilon) \\ & \leq \frac{E(B)}{\varepsilon} = \frac{E(K(f_\rho^{(p)}),\lambda)}{\varepsilon}. \end{aligned}$$

Therefore,

$$\begin{aligned} & Prob^m((z_1,z_2,\cdots,z_m)\in Z^m : B\leq\varepsilon) \\ & \geq 1 - \frac{E(K(f_\rho^{(p)}),\lambda)}{\varepsilon}. \end{aligned} \tag{51}$$

Take $\frac{E(K(f_\rho^{(p)}),\lambda)}{\varepsilon} = \delta$. Then, $\varepsilon = \frac{E(K(f_\rho^{(p)}),\lambda)}{\delta}$. It follows by (51) that, with confidence $1-\delta$, holds

$$B \leq \frac{E(K(f_\rho^{(p)}),\lambda)}{\delta}. \tag{52}$$

(52),(39) and Proposition 10 give (18).

We now show (19). Also, by (14),(12) and the definition of $f_\rho^{(p)}$ we have

$$\begin{aligned} & \mathcal{E}_\rho^{(p)}(f_{\alpha^*_{z,\lambda,p}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & k\sqrt{m}\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\| \\ & +\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}})^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ \leq\ & k\sqrt{m}\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\| + \Big(\mathcal{E}_\rho^{(p)}(f_{\alpha_p^{(\rho),*}}) \\ & +\lambda\|\alpha_p^{(\rho),*}\|^2\Big)^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\ =\ & k\sqrt{m}\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\| \\ & +\Big(\inf_{\alpha\in R^m}(\mathcal{E}_\rho^{(p)}(f_\alpha) + \lambda\|\alpha\|^2)\Big)^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p),*})^{\frac{1}{p}} \\ =\ & k\sqrt{m}\|\alpha^*_{z,\lambda,p} - \alpha_p^{(\rho),*}\| + K^*(f_\rho^{(p)},\lambda) \\ =\ & C + D. \end{aligned} \tag{53}$$

Same way as the proof of (52) shows that, with confidence $1-\delta$, holds

$$D \leq \frac{E(K^*(f_\rho^{(p)}),\lambda)}{\delta} \tag{54}$$

and by (44) we have

$$\begin{aligned} C\ &\leq\ k\sqrt{m}\|\alpha^{(\rho),*}_{z,\lambda,p} - \alpha_p^{(\rho),*}\| \\ &\leq \Big(\frac{c_p}{\delta}\Big)^{\frac{p-1}{p}} \frac{k^2 p|\rho|_p}{\lambda m^{q^*}}. \end{aligned} \tag{55}$$

(54),(53),(55) and Proposition 10 yield (19).

Proof of (20) and (21). Let $L_K^{\frac{1}{2}}$ be the linear operator on $L^2(\rho_X)$ satisfying $L_K^{\frac{1}{2}}\circ L_K^{\frac{1}{2}} = L_K$. Then, $\mathcal{H}_K = L_K^{\frac{1}{2}}(L^2(\rho_X))$ is the usual reproducing kernel Hilbert space [1, 2, 19]. Let $\{\phi_n(x)\}_{n=1}^{+\infty}$ be an orthonormal basis of $L^2(\rho_X)$ which are the eigenfunctions of operator L_K corresponding to eigenvalues $\{\lambda_n\}_{n=1}^{+\infty}$. Then, by [5] we know

$$\begin{aligned} & \mathcal{H}_K \\ =\ & \Big\{f(x) = \sum_{n=1}^\infty \sqrt{\lambda_n}c_n\phi_n(x) \mid \sum_{n=1}^\infty c_n^2 < +\infty\Big\}. \end{aligned}$$

On the other hand,the Mercer theorem (see [6]) gives

$$K(x,y) = \sum_{n=1}^{+\infty}\lambda_n\phi_n(x)\phi_n(y), \qquad x,y\in X. \tag{56}$$

It follows that

$$\begin{aligned} & L_K(L^2(\rho_X)) \\ = & \left\{ f(x) = \sum_{n=1}^{\infty} \lambda_n c_n \phi_n(x) \mid \sum_{n=1}^{\infty} c_n^2 < +\infty \right\} \end{aligned}$$

and $L_K(L^2(\rho_X)) \subset \mathcal{H}_K$. Choose $\alpha^* = \{\alpha_k^*\}_{k=1}^m = \{\frac{\varphi(x_k)}{m}\}_{k=1}^m$. Then, by (14) we know if $f(x) = L_K(\varphi, x)$, then,

$$\begin{aligned} & \Big(\int_X \Big| \sum_{k=1}^m \alpha_k^* K(x_k, x) - f(x) \Big|^p d\rho_X(x) \Big)^{\frac{1}{p}} \\ \leq & k \| \frac{1}{m} \sum_{k=1}^m \varphi(x_k) K(x_k, \cdot) \\ & - \int_X \varphi(u) K(u, \cdot) d\rho_X(u) \|_{K, \overline{X}}. \end{aligned} \tag{57}$$

It follows by $\langle K_x, K_y \rangle_K = K(x,y)$ and (57) that

$$\begin{aligned} & \Big\| \sum_{k=1}^m \alpha_k^* K(x_k, \cdot) - f(\cdot) \Big\|_{L^p(\rho_X)}^2 \\ \leq k^2 & \Big\| \frac{1}{m} \sum_{k=1}^m \varphi(x_k) K(x_k, \cdot) \\ & - \int_X \varphi(u) K(u, \cdot) d\rho_X(u) \Big\|_{K, \overline{X}}^2 \\ = & \frac{k^2}{m^2} \sum_{k=1}^m \sum_{j=1}^m \varphi(x_k) \varphi(x_j) K(x_k, x_j) \\ & - \frac{2k^2}{m} \sum_{k=1}^m \varphi(x_k) \int_X K(x_k, v) \varphi(v) d\rho_X(v) \\ & + k^2 \int_X \int_X \varphi(u) \varphi(v) K(u, v) \; d\rho_X(u) d\rho_X(v). \end{aligned}$$

Therefore, by the Markov inequality we have

$$\begin{aligned} & E\Big(\int_X \Big| \sum_{k=1}^m \alpha_k^* K(x_k, x) - f(x) \Big|^p d\rho_X(x) \Big)^{\frac{1}{p}} \\ \leq & kE\Big(\Big\| \frac{1}{m} \sum_{k=1}^m \varphi(x_k) K(x_k, \cdot) \\ & - \int_X \varphi(u) K(u, \cdot) d\rho_X(u) \Big\|_{K, \overline{X}}^2 \Big)^{\frac{1}{2}} \\ = & k\Big(E\Big[\frac{1}{m^2} \sum_{k=1}^m \sum_{j=1}^m \varphi(x_k) \varphi(x_j) K(x_k, x_j) \\ & - \frac{2}{m} \sum_{k=1}^m \varphi(x_k) \int_X K(x_k, v) \varphi(v) d\rho_X(v) \\ & + \int_X \int_X \varphi(u) \varphi(v) K(u, v) d\rho_X(u) d\rho_X(v) \Big] \Big)^{\frac{1}{2}} \\ = & k\Big(\frac{1}{m^2} \int_{X^m} (\sum_{k=1}^m \sum_{j=1}^m \varphi(x_k) \varphi(x_j) K(x_k, x_j)) \\ & \times d\rho_X(x_1) \cdots d\rho_X(x_m) \\ & - \frac{2}{m} \int_{X^m} (\sum_{k=1}^m \varphi(x_k) \int_X K(x_k, v) \varphi(v) d\rho_X(v)) \\ & \times d\rho_X(x_1) \cdots d\rho_X(x_m) \\ & + \int_X \int_X \varphi(u) \varphi(v) K(u, v) d\rho_X(u) d\rho_X(v) \Big)^{\frac{1}{2}} \\ = & k\Big(- \int_X \int_X \varphi(u) \varphi(v) K(u, v) d\rho_X(u) d\rho_X(v) \\ & + \frac{1}{m} \int_X \varphi(y)^2 K(y, y) d\rho_X(y) \\ & + \frac{m-1}{m} \int_X \int_X \varphi(x) \varphi(y) K(x, y) \\ & \times d\rho_X(x) d\rho_X(y) \Big)^{\frac{1}{2}} \\ = & k \sqrt{\frac{\nu - \mu}{m}}. \end{aligned} \tag{58}$$

Further, the Markov inequality gives

$$\begin{aligned} & E(\|f_{\alpha^*}\|_{K, \overline{X}}^{\frac{1}{p}}) \leq \Big(E(\|f_{\alpha^*}\|_{K, \overline{X}}^2) \Big)^{\frac{1}{2p}} \\ \leq & \Big(E(\frac{1}{m^2} \sum_{i,j=1}^m \varphi(x_i) K(x_i, x_j) \varphi(x_j)) \Big)^{\frac{1}{2p}} \\ = & \Big(\frac{1}{m^2} \int_{X^m} (\sum_{i,j=1, i \neq j}^m \varphi(x_i) K(x_i, x_j) \varphi(x_j)) \\ & \times d\rho_X(x_1) \cdots d\rho_X(x_m) \\ & + \frac{1}{m^2} \int_X (\sum_{i=1}^m \varphi(x_i)^2) d\rho_X(x_1) \cdots d\rho_X(x_m) \Big)^{\frac{1}{2p}} \\ = & \Big(\frac{m-1}{m} \int_X \int_X \varphi(x) K(x, y) \varphi(y) d\rho_X(x) d\rho_X(y) \\ & + \frac{1}{m} \int_X \varphi(x)^2 K(x, x) d\rho_X(x) \Big)^{\frac{1}{2p}} \\ \leq & \Big(\int_X \varphi(x)^2 K(x, x) d\rho_X(x) \Big)^{\frac{1}{2p}} = \sqrt[2p]{\nu}, \end{aligned} \tag{59}$$

where we have used the inequality $K(x,y) \leq \sqrt{K(x,x)K(y,y)}$. Same way yields

$$E(\|\alpha^*\|^{\frac{1}{p}}) \leq \Big(\frac{\|\varphi\|_{L^2(\rho_X)}^2}{m} \Big)^{\frac{1}{2p}}. \tag{60}$$

We now show (20). By (12),(58),(59) and the fact $\mathcal{H}_K \subset L^p(\rho_X)$ we have

$$\begin{aligned} & K(f_\rho^{(p)}, \lambda) \\ \leq & \Big[\mathcal{E}_\rho^{(p)}(f_{\alpha^*}) + \lambda \|f_{\alpha^*}\|_{K, \overline{X}}^2 \Big]^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \end{aligned}$$

$$\begin{aligned}
&\leq \mathcal{E}_\rho^{(p)}(f_{\alpha^*})^{\frac{1}{p}} + \lambda^{\frac{1}{p}}\|f_{\alpha^*}\|_{K,\overline{X}}^{\frac{2}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\
&\leq \|f_{\alpha^*} - f_\rho^{(p)}\|_{L^p(\rho_X)} + \lambda^{\frac{1}{p}}\|f_{\alpha^*}\|_{K,\overline{X}}^{\frac{2}{p}}. \quad (61)
\end{aligned}$$

By (58),(59) and (61) we have

$$E(K(f_\rho^{(p)},\lambda)) \leq \frac{k\sqrt{\nu-\mu}}{\sqrt{m}} + \sqrt[p]{\lambda\nu}. \quad (62)$$

(62) and (18) yield (20).

We now show (21).By (12),(58),(60) and the fact $\mathcal{H}_K \subset L^p(\rho_X)$ we have

$$\begin{aligned}
&K^*(f_\rho^{(p)},\lambda) \\
&\leq \left[\mathcal{E}_\rho^{(p)}(f_{\alpha^*}) + m\lambda\|\alpha^*\|^2\right]^{\frac{1}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\
&\leq \mathcal{E}_\rho^{(p)}(f_{\alpha^*})^{\frac{1}{p}} + m^{\frac{1}{p}}\lambda^{\frac{1}{p}}\|\alpha^*\|^{\frac{2}{p}} - \mathcal{E}_\rho^{(p)}(f_\rho^{(p)})^{\frac{1}{p}} \\
&\leq \|f_{\alpha^*} - f_\rho^{(p)}\|_{L^p(\rho_X)} + (m\lambda)^{\frac{1}{p}}\|\alpha^*\|^{\frac{2}{p}}. \quad (63)
\end{aligned}$$

By (60),(58)and (63) we have

$$\begin{aligned}
&E(K^*(f_\rho^{(p)},\lambda)) \\
&\leq \frac{k\sqrt{\nu-\mu}}{\sqrt{m}} + \sqrt[p]{\lambda\|\varphi\|_{L^2(\rho_X)}}. \quad (64)
\end{aligned}$$

Therefore, (64)and (19) yield (21).

Acknowledgements: The work is supported by the NSF of China under grant numbers: 10871226, 10971251, 11101220 11271199 and the NSF of Zhejiang Province under grant numbers: Y6100096 and LQ12F02007

References:

[1] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.*, vol.68,1950, no. 2, pp. 337-404

[2] C. Carmeli, E. De Vito, A. Toigo, Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem, *Anal. and Appli.*, vol.4, 2006, no. 4, pp. 377-408

[3] A. Christmann, I. Steinwart, Consistency and robustness of kernel-based regression in convex risk minimization, *Bernoulli*, vol.13,2007,no.3,pp. 799-819

[4] A. Christmann, I. Steinwart, Consistency of kernel based quantile regression, *Appl. Stoch. Model. Bus. and Industr.*, vol.24, 2008, no.2, pp.171-183

[5] F. Cucker, S. Smale, On the mathematical foundations of learning theory, *Bull. Amer. Math.Soc.*, vol.39, 2001,no.1, pp. 1-49

[6] F. Cucker, D. X. Zhou, *Learning Theory: An Approximation Theory Viewpoint,* Cambridge University Press, New York, 2007

[7] C. De Mol, E. De Vito, L. Rosasco, Elastic-net regularization in learning theory, *J. Complexity*, vol.25, 2009, no.2, pp. 201-230

[8] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, *Adv. Comput. Math.*, vol.13,2000,no.1, pp.1-50

[9] J.-B.Hiriart-Urruty, C. Lemaréchas, *Fundamental of Convex Analysis*, Springer-Verlag, Berlin, 2001

[10] B. H. Sheng, Z. X. Chen, J. L. Wang, Learning rates of Tikhonov regularized regressions based on sample dependent RKHS, *J. Comput. Anal. and Appl.*,vol.14, 2012, no.2, pp. 145-163

[11] B. H. Sheng, P. X. Ye, Error analysis for support vector machine classifiers on unit sphere of Euclidean space,*J. Comput. Inform. Syst.*, vol.7,2011, no.9, pp. 3023-3030

[12] B. H. Sheng,P. X. Ye, Least square regression learning with data dependent hypothesis and coefficient regularization, *J. Computer*, vol.6, 2011, no.4, pp. 671-675

[13] B. H. Sheng, D. H. Xiang, The consistency analysis of coefficient regularized classification with convex loss, *WSEAS Trans. on Math.*, vol.10, 2011, no.9, pp. 291-300

[14] B.H.Sheng, D. H. Xiang, Bound the learning rates with generalized gradients, *WSEAS Trans. on Signal Processing*, vol. 8, 2012, no.1, pp.1-10

[15] B. H. Sheng, P. X. Ye, J. L.Wang, Learning rates for least square regressions with coefficient regularization, *Acta Math. Sinica, English Series*, vol. 28, 2012, no.11, pp. 2205-2212

[16] B. H. Sheng, P. X. Ye, Z. X. Chen, *The Nonsmooth Analysis Method in Kernel Learning*, Science Press, Beijing, 2012

[17] L. Shi,Y. L. Feng, D. X. Zhou, Concentraction estimates for learning with l^1-regularizer and data dependent hypothesis spaces, *Applied and Comput. Harm. Anal.*, vol. 31, 2011, no.2, pp. 286-302

[18] S. Smale, D. X. Zhou, Learning theory estimates via integral operators and their applications, *Constr. Approx.*, vol.26,2007,no.2, pp.153-172

[19] S. Smale, D. X.Zhou, Estimating the approximation error in learning theory, *Anal. and Appl.*vol.1, 2003, no.1, pp. 17-41

[20] S. Smale, D. X. Zhou, Shannon sampling and function reconstruction from point values, *Bull. (New Series) Amer. Math. Soc.*, vol.41, 2004, no.3, pp.279-305

[21] H. W. Sun, Q. Wu, Least square regression with indefinite kernels and coefficient regularization, *Applied Comput. Harm. Anal.*, vol.30,2011, no.1, pp. 96-109

[22] H. Z. Tong, D. R. Chen, F. H. Yang, Least square regression with l^p-coefficient regularization, *Neural Comput.*, vol.22, 2010, no.12, pp.3221-3235

[23] C. Wang, Z. C. Guo, ERM learning with unbounded sampling,*Acta Math. Sinica, English Series*,vol.28, 2012, no.11, pp. 97-104

[24] Y. Y. Wang, Y. Zhou, Z. Chen, S. S.Gao, B. H. Sheng, The convergence rate of learning algorithm for least square regression with sample dependent hypothesis spaces,*J. Baoji University of Arts and Sciences(Natural Science)*, vol.31,2011,no.3, pp.5-9

[25] Q. Wu, D.X. Zhou, Learning with sample dependent hypothesis spaces, *Comput. and Math. with Appli.*, vol.56,2008, no.11,pp. 2896-2907

[26] Q. W. Xiao, D. X. Zhou, Learning by nonsymmetric kernels with data dependent spaces and l^1-regularizer,*Taiwanese J. Math.*, vol.14, 2010, no.5, pp. 1821-1836

[27] J. Zhang, J. L. Wang, B. H. Sheng, Learning from regularized regression algorithms with p-order Markov chain sampling, *Appl. Math. J. Chinese Univ.*, vol.26, 2011, no.3,pp.295-306.