

Research Article

Approximation Analysis of Learning Algorithms for Support Vector Regression and Quantile Regression

Dao-Hong Xiang,¹ Ting Hu,² and Ding-Xuan Zhou³

¹ Department of Mathematics, Zhejiang Normal University, Zhejiang 321004, China

² School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

³ Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

Correspondence should be addressed to Ding-Xuan Zhou, mazhou@cityu.edu.hk

Received 14 July 2011; Accepted 14 November 2011

Academic Editor: Yuesheng Xu

Copyright © 2012 Dao-Hong Xiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study learning algorithms generated by regularization schemes in reproducing kernel Hilbert spaces associated with an ϵ -insensitive pinball loss. This loss function is motivated by the ϵ -insensitive loss for support vector regression and the pinball loss for quantile regression. Approximation analysis is conducted for these algorithms by means of a variance-expectation bound when a noise condition is satisfied for the underlying probability measure. The rates are explicitly derived under a priori conditions on approximation and capacity of the reproducing kernel Hilbert space. As an application, we get approximation orders for the support vector regression and the quantile regularized regression.

1. Introduction and Motivation

In this paper, we study a family of learning algorithms serving both purposes of *support vector regression* and *quantile regression*. Approximation analysis and learning rates will be provided, which also helps better understanding of some classical learning methods.

Support vector regression is a classical kernel-based algorithm in learning theory introduced in [1]. It is a regularization scheme in a *reproducing kernel Hilbert space* (RKHS) \mathcal{H}_K associated with an ϵ -insensitive loss $\psi^\epsilon : \mathbb{R} \rightarrow \mathbb{R}_+$ defined for $\epsilon \geq 0$ by

$$\psi^\epsilon(u) = \max\{|u| - \epsilon, 0\} = \begin{cases} |u| - \epsilon, & \text{if } |u| \geq \epsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (1.1)$$

Here, for learning functions on a compact metric space X , $K : X \times X \rightarrow \mathbb{R}$ is a continuous, symmetric, and positive semidefinite function called a *Mercer kernel*. The associated RKHS \mathcal{H}_K is defined [2] as the completion of the linear span of the set of function $\{K_x = K(x, \cdot) : x \in X\}$ with the inner product $\langle \cdot, \cdot \rangle_K$ satisfying $\langle K_x, K_y \rangle_K = K(x, y)$. Let $Y = \mathbb{R}$ and ρ be a Borel probability measure on $Z := X \times Y$. With a sample $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m \in Z^m$ independently drawn according to ρ , the support vector regression is defined as

$$f_{\mathbf{z}}^{\text{SVR}} = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m \psi^\epsilon(f(x_i) - y_i) + \lambda \|f\|_K^2 \right\}, \quad (1.2)$$

where $\lambda = \lambda(m) > 0$ is a *regularization parameter*.

When $\epsilon > 0$ is fixed, convergence of (1.2) was analyzed in [3]. Notice from the original motivation [1] for the insensitive parameter ϵ for balancing the approximation and sparsity of the algorithm that ϵ should change with the sample size and usually $\epsilon = \epsilon(m) \rightarrow 0$ as the sample size m increases. Mathematical analysis for this original algorithm is still open. We will solve this problem in a special case of our approximation analysis for general learning algorithms. In particular, we show how $f_{\mathbf{z}}^{\text{SVR}}$ approximates the median function $f_{\rho, 1/2}$, which is one of the purposes of this paper. Here, for $x \in X$, the median function value $f_{\rho, 1/2}(x)$ is a median of the conditional distribution $\rho(\cdot | x)$ of ρ at x .

Quantile regression, compared with the least squares regression, provides richer information about response variables such as stretching or compressing tails [4]. It aims at estimating quantile regression functions. With a quantile parameter $0 < \tau < 1$, a *quantile regression function* $f_{\rho, \tau}$ is defined by its value $f_{\rho, \tau}(x)$ to be a τ -quantile of $\rho(\cdot | x)$, that is, a value $u \in Y$ satisfying

$$\rho(\{y \in Y : y \leq u\} | x) \geq \tau, \quad \rho(\{y \in Y : y \geq u\} | x) \geq 1 - \tau. \quad (1.3)$$

Quantile regression has been studied by kernel-based regularization schemes in a learning theory literature (e.g., [5–8]). These regularization schemes take the form

$$f_{\mathbf{z}}^{\text{QR}} = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m \psi_\tau(f(x_i) - y_i) + \lambda \|f\|_K^2 \right\}, \quad (1.4)$$

where $\psi_\tau : \mathbb{R} \rightarrow \mathbb{R}_+$ is the pinball loss shown in Figure 1 defined by

$$\psi_\tau(u) = \begin{cases} (1 - \tau)u, & \text{if } u > 0, \\ -\tau u, & \text{if } u \leq 0. \end{cases} \quad (1.5)$$

Motivated by the ϵ -insensitive loss ψ^ϵ and the pinball loss ψ_τ , we propose the ϵ -insensitive pinball loss $\psi_\tau^\epsilon : \mathbb{R} \rightarrow \mathbb{R}_+$ with an insensitive parameter $\epsilon \geq 0$ shown in Figure 1 defined as

$$\psi_\tau^\epsilon(u) = \begin{cases} (1 - \tau)(u - \epsilon), & \text{if } u > \epsilon, \\ -\tau(u + \epsilon), & \text{if } u \leq -\epsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (1.6)$$

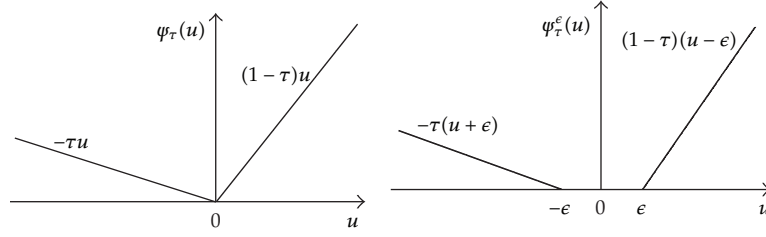


Figure 1

This loss function has been applied to online learning for quantile regression in our previous work [8]. It is applied here to a regularization scheme in the RKHS as

$$f_z^{(\epsilon)} = f_{z,\lambda,\tau}^{(\epsilon)} = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^m \psi_\tau^\epsilon(f(x_i) - y_i) + \lambda \|f\|_K^2 \right\}. \quad (1.7)$$

The main goal of this paper is to study how the output function $f_z^{(\epsilon)}$ given by (1.7) converges to the quantile regression function $f_{\rho,\tau}$ and how explicit learning rates can be obtained with suitable choices of the parameters $\lambda = m^{-\alpha}$, $\epsilon = m^{-\beta}$ based on a priori conditions on the probability measure ρ .

2. Main Results on Approximation

Throughout the paper, we assume that the conditional distribution $\rho(\cdot | x)$ is supported on $[-1, 1]$ for every $x \in X$. Then, we see from (1.3) that we can take values of $f_{\rho,\tau}$ to be on $[-1, 1]$. So to see how $f_z^{(\epsilon)}$ approximates $f_{\rho,\tau}$, it is natural to project values of the output function $f_z^{(\epsilon)}$ onto the same interval by the projection operator introduced in [9].

Definition 2.1. The projection operator π on the space of function on X is defined by

$$\pi(f)(x) = \begin{cases} 1, & \text{if } f(x) > 1, \\ -1, & \text{if } f(x) < -1, \\ f(x), & \text{if } -1 \leq f(x) \leq 1. \end{cases} \quad (2.1)$$

Our approximation analysis aims at establishing bounds for the error $\|\pi(f_z^{(\epsilon)}) - f_{\rho,\tau}\|_{L_{\rho_X}^{p^*}}$ in the space $L_{\rho_X}^{p^*}$ with some $p^* > 0$ where ρ_X is the marginal distribution of ρ on X .

2.1. Support Vector Regression and Quantile Regression

Our error bounds and learning rates are presented in terms of a noise condition and approximation condition on ρ .

The noise condition on ρ is defined in [5, 6] as follows.

Definition 2.2. Let $p \in (0, \infty]$ and $q \in (1, \infty)$. We say that ρ has a τ -quantile of p -average type q if for every $x \in X$, there exist a τ -quantile $t^* \in \mathbb{R}$ and constants $a_x \in (0, 2]$, $b_x > 0$ such that for each $u \in [0, a_x]$,

$$\rho(\{y \in (t^* - u, t^*)\} | x) \geq b_x u^{q-1}, \quad \rho(\{y \in (t^*, t^* + u)\} | x) \geq b_x u^{q-1}, \quad (2.2)$$

and that the function on X taking value $(b_x a_x^{q-1})^{-1}$ at $x \in X$ lies in $L_{\rho_x}^p$.

Note that condition (2.2) tells us that $f_{\rho, \tau}(x) = t^*$ is uniquely defined at every $x \in X$.

The approximation condition on ρ is stated in terms of the integral operator $L_K : L_{\rho_x}^2 \rightarrow L_{\rho_x}^2$ defined by $L_K(f)(x) = \int_X K(x, u) f(u) d\rho_X(u)$. Since K is positive semidefinite, L_K is a compact positive operator and its r -th power L_K^r is well-defined for any $r > 0$. Our approximation condition is given as

$$f_{\rho, \tau} = L_K^r(g_{\rho, \tau}) \quad \text{for some } 0 < r \leq \frac{1}{2}, \quad g_{\rho, \tau} \in L_{\rho_x}^2. \quad (2.3)$$

Let us illustrate our approximation analysis by the following special case which will be proved in Section 5.

Theorem 2.3. *Let $X \subset \mathbb{R}^n$ and $K \in C^\infty(X \times X)$. Assume $f_{\rho, \tau} \in \mathcal{H}_K$ and ρ has a τ -quantile of p -average type 2 for some $p \in (0, \infty]$. Take $\lambda = m^{-(p+1)/(p+2)}$ and $\epsilon = m^{-\beta}$ with $(p+1)/(p+2) \leq \beta \leq \infty$. Let $0 < \eta < (p+1)/2(p+2)$. Then, with $p^* = 2p/(p+1) > 0$, for any $0 < \delta < 1$, with confidence $1 - \delta$, one has*

$$\left\| \pi\left(f_z^{(\epsilon)}\right) - f_{\rho, \tau} \right\|_{L_{\rho_x}^{p^*}} \leq \tilde{C} \log \frac{3}{\delta} m^{\eta - (p+1)/2(p+2)}, \quad (2.4)$$

where \tilde{C} is a constant independent of m or δ .

If $p \geq (1/2\eta) - 2$ for $0 < \eta < 1/4$, we see that the power exponent for the learning rate (2.4) is at least $(1/2) - 2\eta$. This exponent can be arbitrarily close to $1/2$ when η is small enough.

In particular, if we take $\tau = 1/2$, Theorem 2.3 provides rates for output function f_z^{SVR} of the support vector regression (1.2) to approximate the median function $f_{\rho, 1/2}$.

If we take $\beta = \infty$ leading to $\epsilon = 0$, Theorem 2.3 provides rates for output function f_z^{QR} of the quantile regression algorithm (1.4) to approximate the quantile regression function $f_{\rho, \tau}$.

2.2. General Approximation Analysis

To state our approximation analysis in the general case, we need the capacity of the hypothesis space measured by covering numbers.

Definition 2.4. For a subset S of $C(X)$ and $u > 0$, the *covering number* $\mathcal{N}(S, u)$ is the minimal integer $l \in \mathbb{N}$ such that there exist l disks with radius u covering S .

The covering numbers of balls $B_R = \{f \in \mathcal{H}_K : \|f\|_K \leq R\}$ with radius $R > 0$ of the RKHS have been well studied in the learning theory literature [10, 11]. In this paper, we assume for some $s > 0$ and $C_s > 0$ that

$$\log \mathcal{N}(B_1, u) \leq C_s \left(\frac{1}{u}\right)^s, \quad \forall u > 0. \quad (2.5)$$

Now we can state our main result which will be proved in Section 5. For $p \in (0, \infty]$ and $q \in (1, \infty)$, we denote

$$\theta = \min\left\{\frac{2}{q}, \frac{p}{p+1}\right\} \in (0, 1]. \quad (2.6)$$

Theorem 2.5. *Assume (2.3) with $0 < r \leq 1/2$ and (2.5) with $s > 0$. Suppose that ρ has a τ -quantile of p -average type q for some $p \in (0, \infty]$ and $q \in (1, \infty)$. Take $\lambda = m^{-\alpha}$ with $0 < \alpha \leq 1$ and $\alpha < (2+s)/s(2+s-\theta)$. Set $\epsilon = m^{-\beta}$ with $\alpha r/(1-r) \leq \beta \leq \infty$. Let*

$$0 < \eta < \frac{(1+s)[2+2s-s\alpha(2+s-\theta)]}{s(2+s-\theta)(2+s)}. \quad (2.7)$$

Then, with $p^* = pq/(p+1) > 0$, for any $0 < \delta < 1$, with confidence $1 - \delta$, one has

$$\left\| \pi(f_z^{(\epsilon)}) - f_{\rho, \tau} \right\|_{L_{\rho_X}^{p^*}} \leq \tilde{C} \left(\log \frac{3}{\eta} \right)^2 \log \frac{3}{\delta} m^{-\vartheta}, \quad (2.8)$$

where \tilde{C} is a constant independent of m or δ and the power index ϑ is given in terms of r, s, p, q, α , and η by

$$\vartheta = \frac{1}{q} \min \left\{ \frac{\alpha r}{1-r}, \frac{1}{2+s-\theta} - \frac{s[\alpha(2+s-\theta)-1]}{(2+s-\theta)(2+s)} - \frac{s\eta}{1+s}, \right. \\ \left. \frac{1}{2+s-\theta} - \frac{s\alpha(1-2r)}{(1+s)(2-2r)}, \frac{1}{2+s-\theta} - \frac{s}{1+s} \left(\frac{\alpha}{2} - \frac{1}{2(2-\theta)} \right) \right\}. \quad (2.9)$$

The index ϑ can be viewed as a function of variables r, s, p, q, α, η . The restriction $0 < \alpha < (2+s)/s(2+s-\theta)$ on α and (2.7) on η ensure that ϑ is positive, which verifies the valid learning rate in Theorem 2.5.

Assumption (2.5) is a measurement of regularity of the kernel K when X is a subset of \mathbb{R}^n . In particular, s can be arbitrarily small when K is smooth enough. In this case, the power index ϑ in (2.8) can be arbitrarily close to $(1/q) \min\{\alpha r/(1-r), 1/(2-\theta)\}$. Again, when $\beta = \infty$, $\epsilon = 0$, algorithm (1.7) corresponds to algorithm (1.4) for quantile regression. In this case, Theorem 2.5 provides learning rates for quantile regression algorithm (1.4).

Error analysis has been done for quantile regression algorithm (1.4) in [5, 6]. Under the assumptions that ρ satisfies (2.2) with some $p \in (0, \infty]$ and $q > 1$ and the ℓ^2 empirical covering number $\mathcal{N}_z(B_1, \eta, \ell^2)$ (see [5] for more details) of B_1 is bounded as

$$\sup_{z \in Z^m} \log \mathcal{N}_z(B_1, \eta, \ell^2) \leq a \left(\frac{1}{\eta} \right)^s \quad \text{with } s \in (0, 2), \quad a \geq 1, \quad (2.10)$$

it was proved in [5] that with confidence $1 - \delta$,

$$\begin{aligned} \left\| \pi(f_z^{\text{QR}}) - f_{\rho, \tau} \right\|_{L_{\rho_X}^{p^*}} \leq & \left\{ \mathfrak{D}_\tau(\lambda) + \sqrt{\frac{\mathfrak{D}_\tau(\lambda) \log(3/\delta)}{\lambda} \frac{1}{m}} + \left(\frac{K_{s, C_p} a}{\lambda^{s/2} m} \right)^{(p+1)/(p+2-s/2)} \right. \\ & \left. + \frac{K_{s, C_p} a}{\lambda^{s/2} m} + 5 \left(\frac{32 C_p \log(3/\delta)}{m} \right)^{(p+1)/(p+2)} + \frac{145 \log(3/\delta)}{m} \right\}^{1/q}, \end{aligned} \quad (2.11)$$

where C_p and K_{s, C_p} are constants independent of m or λ . Here, $\mathfrak{D}_\tau(\lambda)$ is the *regularization error* defined as

$$\mathfrak{D}_\tau(\lambda) = \inf_{f \in \mathcal{L}_K} \left\{ \mathcal{E}_\tau(f) - \mathcal{E}_\tau(f_{\rho, \tau}) + \lambda \|f\|_K^2 \right\}, \quad (2.12)$$

and $\mathcal{E}_\tau(f)$ is the generalization error associated with the pinball loss ψ_τ defined by

$$\mathcal{E}_\tau(f) = \int_Z \psi_\tau(f(x) - y) d\rho = \int_X \int_Y \psi_\tau(f(x) - y) d\rho(y | x) d\rho_X(x). \quad (2.13)$$

Note that $\mathcal{E}_\tau(f)$ is minimized by the quantile regression function $f_{\rho, \tau}$. Thus, when the regularization error $\mathfrak{D}_\tau(\lambda)$ decays polynomially as $\mathfrak{D}_\tau(\lambda) = O(\lambda^{r/(1-r)})$ (which is ensured by Lemma 2.6 below when (2.3) is satisfied) and $\lambda = m^{-\alpha}$, then $\left\| \pi(f_z^{\text{QR}}) - f_{\rho, \tau} \right\|_{L_{\rho_X}^{p^*}} = O(\log(3/\delta) m^{-\vartheta})$ with

$$\vartheta = \frac{1}{q} \min \left\{ \frac{\alpha r}{1-r}, \frac{p+1}{p+2-(s/2)} \left(1 - \frac{\alpha s}{2} \right), \frac{p+1}{p+2} \right\}. \quad (2.14)$$

Since $(p+1)/(p+2) = 1/(2-\theta)$, we see that this learning rate is comparable to our result in (2.8).

2.3. Comparison with Least Squares Regression

There has been a large literature in learning theory (described in [12]) for the least squares algorithms:

$$f_z^{LS} = \arg \min_{f \in \mathcal{L}_K} \left\{ \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 + \lambda \|f\|_K^2 \right\}. \quad (2.15)$$

It aims at learning the regression function $f_\rho(x) = \int_Y y d\rho(y | x)$. A crucial property for its error analysis is the identity $\mathcal{E}_{ls}(f) - \mathcal{E}_{ls}(f_\rho) = \|f - f_\rho\|_{L_{\rho_X}^2}^2$ for the least squares generalization error $\mathcal{E}_{ls}(f) = \int_Z (y - f(x))^2 d\rho$. It yields a variance-expectation bound $\mathbf{E}(\xi^2) \leq 4\mathbf{E}(\xi)$ for the random variable $\xi = (y - f(x))^2 - (y - f_\rho(x))^2$ on (Z, ρ) where $f : X \rightarrow Y$ is an arbitrary measurable function. Such a variance-expectation bound with $\mathbf{E}(\xi)$ possibly replaced by its positive power $(\mathbf{E}(\xi))^\theta$ plays an essential role for analyzing regularization schemes and the power exponent θ depends on strong convexity of the loss. See [13] and references therein. However, the pinball loss in the quantile regression setting has no strong convexity [6] and we would not expect a variance-expectation bound for a general distribution ρ . When ρ has a τ -quantile of p -average type q , the following variance-expectation bound with θ given by (2.6) can be found in [5, 7] (derived by means of Lemma 3.1 below).

Lemma 2.6. *If ρ has a τ -quantile of p -average type q for some $p \in (0, \infty]$ and $q \in (1, \infty)$, then*

$$\mathbf{E} \left\{ (\psi_\tau(f(x) - y) - \psi_\tau(f_{\rho,\tau}(x) - y))^2 \right\} \leq C_\theta (\mathcal{E}_\tau(f) - \mathcal{E}_\tau(f_{\rho,\tau}))^\theta, \quad \forall f : X \rightarrow Y, \quad (2.16)$$

where the power index θ is given by (2.6) and the constant C_θ is $C_\theta = 2^{2-\theta} q^\theta \|\gamma^{-1}\|_{L_{\rho_X}^p}^\theta$.

Lemma 2.6 overcomes the difficulty of quantile regression caused by lack of strong convexity of the pinball loss. It enables us to derive satisfactory learning rates, as in Theorem 2.5.

3. Insensitive Relation and Error Decomposition

An important relation for quantile regression observed in [5] asserts that the error $\|\pi(f_z^{(e)}) - f_{\rho,\tau}\|$ taken in a suitable $L_{\rho_X}^{p^*}$ space can be bounded by the excess generalization error $\mathcal{E}_\tau(\pi(f_z^{(e)})) - \mathcal{E}_\tau(f_{\rho,\tau})$ when the noise condition is satisfied.

Lemma 3.1. *Let $p \in (0, \infty]$ and $q \in (1, \infty)$. Denote $p^* = pq/(p+1) > 0$. If ρ has a τ -quantile of p -average type q , then for any measurable function f on X , one has*

$$\|f - f_{\rho,\tau}\|_{L_{\rho_X}^{p^*}} \leq C_{q,\rho} \{ \mathcal{E}_\tau(f) - \mathcal{E}_\tau(f_{\rho,\tau}) \}^{1/q}, \quad (3.1)$$

where $C_{q,\rho} = 2^{1-(1/q)} q^{1/q} \| \{ (b_x a_x^{q-1})^{-1} \}_{x \in X} \|_{L_{\rho_X}^p}^{1/q}$.

By Lemma 3.1, to estimate $\|\pi(f_z^{(\epsilon)}) - f_{\rho,\tau}\|_{L_{\rho X}^p}$, we only need to bound the excess generalization error $\mathcal{E}_\tau(\pi(f_z^{(\epsilon)})) - \mathcal{E}_\tau(f_{\rho,\tau})$. This will be done by conducting an error decomposition which has been developed in the literature for regularization schemes [9, 13–15]. Technical difficulty arises for our problem here because the insensitive parameter ϵ changes with m . This can be overcome [16] by the following insensitive relation

$$\psi_\tau(u) - \epsilon \leq \psi_\tau^\epsilon(u) \leq \psi_\tau(u), \quad u \in \mathbb{R}. \quad (3.2)$$

Now, we can conduct an error decomposition. Define the empirical error $\mathcal{E}_{z,\tau}(f)$ for $f : X \rightarrow \mathbb{R}$ as

$$\mathcal{E}_{z,\tau}(f) = \frac{1}{m} \sum_{i=1}^m \psi_\tau(f(x_i) - y_i). \quad (3.3)$$

Lemma 3.2. Let $\lambda > 0$, $f_z^{(\epsilon)}$ be defined by (1.7) and

$$f_\lambda^{(0)} = \arg \min_{f \in \mathcal{L}_K} \left\{ \mathcal{E}_\tau(f) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \|f\|_K^2 \right\}. \quad (3.4)$$

Then,

$$\mathcal{E}_\tau(\pi(f_z^{(\epsilon)})) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \|f_z^{(\epsilon)}\|_K^2 \leq \mathcal{S}_1 + \mathcal{S}_2 + \epsilon + \mathfrak{D}_\tau(\lambda), \quad (3.5)$$

where

$$\begin{aligned} \mathcal{S}_1 &= \left[\mathcal{E}_\tau(\pi(f_z^{(\epsilon)})) - \mathcal{E}_\tau(f_{\rho,\tau}) \right] - \left[\mathcal{E}_{z,\tau}(\pi(f_z^{(\epsilon)})) - \mathcal{E}_{z,\tau}(f_{\rho,\tau}) \right], \\ \mathcal{S}_2 &= \left[\mathcal{E}_{z,\tau}(f_\lambda^{(0)}) - \mathcal{E}_{z,\tau}(f_{\rho,\tau}) \right] - \left[\mathcal{E}_\tau(f_\lambda^{(0)}) - \mathcal{E}_\tau(f_{\rho,\tau}) \right], \end{aligned} \quad (3.6)$$

Proof. The regularized excess generalization error $\mathcal{E}_\tau(\pi(f_z^{(\epsilon)})) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \|f_z^{(\epsilon)}\|_K^2$ can be expressed as

$$\begin{aligned} & \left\{ \mathcal{E}_\tau(\pi(f_z^{(\epsilon)})) - \mathcal{E}_{z,\tau}(\pi(f_z^{(\epsilon)})) \right\} + \left\{ \left[\mathcal{E}_{z,\tau}(\pi(f_z^{(\epsilon)})) + \lambda \|f_z^{(\epsilon)}\|_K^2 \right] - \left[\mathcal{E}_{z,\tau}(f_\lambda^{(0)}) + \lambda \|f_\lambda^{(0)}\|_K^2 \right] \right\} \\ & + \left\{ \left[\mathcal{E}_{z,\tau}(f_\lambda^{(0)}) - \mathcal{E}_\tau(f_\lambda^{(0)}) \right] \right\} + \left\{ \mathcal{E}_\tau(f_\lambda^{(0)}) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \|f_\lambda^{(0)}\|_K^2 \right\}. \end{aligned} \quad (3.7)$$

The fact $|y| \leq 1$ implies that $\mathcal{E}_{z,\tau}(f_z^{(e)}) \leq \mathcal{E}_{z,\tau}(f_z^{(e)})$. The insensitive relation (3.2) and the definition of $f_z^{(e)}$ tell us that

$$\begin{aligned} \mathcal{E}_{z,\tau}(f_z^{(e)}) + \lambda \|f_z^{(e)}\|_K^2 &\leq \frac{1}{m} \sum_{i=1}^m \psi_\tau^\varepsilon(f_z^{(e)}(x_i) - y_i) + \lambda \|f_z^{(e)}\|_K^2 + \varepsilon \\ &\leq \frac{1}{m} \sum_{i=1}^m \psi_\tau^\varepsilon(f_\lambda^{(0)}(x_i) - y_i) + \lambda \|f_\lambda^{(0)}\|_K^2 + \varepsilon \leq \mathcal{E}_{z,\tau}(f_\lambda^{(0)}) + \lambda \|f_\lambda^{(0)}\|_K^2 + \varepsilon. \end{aligned} \quad (3.8)$$

Then, by subtracting and adding $\mathcal{E}_\tau(f_{\rho,\tau})$ and $\mathcal{E}_{z,\tau}(f_{\rho,\tau})$ and noting $\mathfrak{D}_\tau(\lambda) = \mathcal{E}_\tau(f_\lambda^{(0)}) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \|f_\lambda^{(0)}\|_K^2$, we see that the desired inequality in Lemma 3.2 holds true. \square

In the error decomposition (3.5), the first two terms are called *sample error*. The last term is the regularization error defined in (2.12). It can be estimated as follows.

Proposition 3.3. *Assume (2.3). Define $f_\lambda^{(0)}$ by (3.4). Then, one has*

$$\begin{aligned} \mathfrak{D}_\tau(\lambda) &\leq C_0 \lambda^{r/(1-r)}, \\ \|f_\lambda^{(0)}\|_K &\leq \sqrt{C_0} \lambda^{(2r-1)/(2-2r)}, \end{aligned} \quad (3.9)$$

where C_0 is the constant $C_0 = \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2} + \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}^2$.

Proof. Let $\mu = \lambda^{1/(1-r)} > 0$ and

$$f_\mu = (L_K + \mu I)^{-1} L_K f_{\rho,\tau}. \quad (3.10)$$

It can be found in [17, 18] that when (2.3) holds, we have

$$\|f_\mu - f_{\rho,\tau}\|_{L_{\rho_X}^2}^2 + \mu \|f_\mu\|_K^2 \leq \mu^{2r} \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}^2. \quad (3.11)$$

Hence, $\|f_\mu - f_{\rho,\tau}\|_{L_{\rho_X}^2} \leq \mu^r \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}$ and $\|f_\mu\|_K^2 \leq \mu^{2r-1} \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}^2$. Since ψ_τ is Lipschitz, we know by taking $f = f_\mu$ in (2.12) that

$$\begin{aligned} \mathfrak{D}_\tau(\lambda) &\leq \int_Z \psi_\tau(f_\mu(x) - y) - \psi_\tau(f_{\rho,\tau}(x) - y) + \lambda \|f_\mu\|_K^2 \\ &\leq \|f_\mu - f_{\rho,\tau}\|_{L_{\rho_X}^1} + \lambda \|f_\mu\|_K^2 \leq \|f_\mu - f_{\rho,\tau}\|_{L_{\rho_X}^2} + \lambda \|f_\mu\|_K^2 \\ &\leq \mu^r \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2} + \lambda \mu^{2r-1} \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}^2 = \left(\|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2} + \|\mathcal{g}_{\rho,\tau}\|_{L_{\rho_X}^2}^2 \right) \lambda^{r/(1-r)}. \end{aligned} \quad (3.12)$$

This verifies the desired bound for $\mathfrak{D}_\tau(\lambda)$. By taking $f = 0$ in (2.12), we have

$$\lambda \|f_\lambda^{(0)}\|_K^2 \leq \mathfrak{D}_\tau(\lambda) \leq C_0 \lambda^{r/(1-r)}. \quad (3.13)$$

Then the bound for $\|f_\lambda^{(0)}\|_K$ is proved. \square

4. Estimating Sample Error

This section is devoted to estimating the sample error. This is conducted by using the variance-expectation bound in Lemma 2.6.

Denote $\kappa = \sup_{x \in X} \sqrt{K(x, x)}$. For $R \geq 1$, denote

$$\mathcal{W}(R) = \left\{ \mathbf{z} \in Z^m : \left\| f_{\mathbf{z}}^{(\varepsilon)} \right\|_K \leq R \right\}. \quad (4.1)$$

Proposition 4.1. *Assume (2.3) and (2.5). Let $R \geq 1$ and $0 < \delta < 1$. If ρ has a τ -quantile of p -average type q for some $p \in (0, \infty]$ and $q \in (1, \infty)$, then there exists a subset V_R of Z^m with measure at most δ such that for any $\mathbf{z} \in \mathcal{W}(R) \setminus V_R$,*

$$\begin{aligned} \mathcal{E}_\tau \left(\pi \left(f_{\mathbf{z}}^{(\varepsilon)} \right) \right) - \mathcal{E}_\tau (f_{\rho, \tau}) + \lambda \left\| f_{\mathbf{z}}^{(\varepsilon)} \right\|_K^2 &\leq 2\varepsilon + C'_1 \log \frac{3}{\delta} \lambda^{r/(1-r)} \max \left\{ 1, \frac{\lambda^{-(1/(2-2r))}}{m} \right\} \\ &\quad + C'_2 \log \frac{3}{\delta} m^{-1/(2-\theta)} + C'_3 m^{-1/(2+s-\theta)} R^{s/(1+s)}, \end{aligned} \quad (4.2)$$

where θ is given by (2.6) and C'_1, C'_2, C'_3 are constants given by

$$C'_1 = 4C_0 + 2\kappa\sqrt{C_0}, \quad C'_2 = 245 + 324C_\theta^{1/(2-\theta)}, \quad C'_3 = 40(6C_s)^{1/(1+s)} + 40(8C_\theta C_s)^{1/(2+s-\theta)}. \quad (4.3)$$

Proof. Let us first estimate the second part \mathcal{S}_2 of the sample error. It can be decomposed into two parts $\mathcal{S}_2 = \mathcal{S}_{2,1} + \mathcal{S}_{2,2}$ where

$$\begin{aligned} \mathcal{S}_{2,1} &= \left[\mathcal{E}_{\mathbf{z}, \xi} \left(f_{\lambda}^{(0)} \right) - \mathcal{E}_{\mathbf{z}, \tau} \left(\pi \left(f_{\lambda}^{(0)} \right) \right) \right] - \left[\mathcal{E}_\tau \left(f_{\lambda}^{(0)} \right) - \mathcal{E}_\tau \left(\pi \left(f_{\lambda}^{(0)} \right) \right) \right], \\ \mathcal{S}_{2,2} &= \left[\mathcal{E}_{\mathbf{z}, \tau} \left(\pi \left(f_{\lambda}^{(0)} \right) \right) - \mathcal{E}_{\mathbf{z}, \tau} (f_{\rho, \tau}) \right] - \left[\mathcal{E}_\tau \left(\pi \left(f_{\lambda}^{(0)} \right) \right) - \mathcal{E}_\tau (f_{\rho, \tau}) \right]. \end{aligned} \quad (4.4)$$

For bounding $\mathcal{S}_{2,1}$, we take the random variable $\xi(\mathbf{z}) = \psi_\tau(f_{\lambda}^{(0)}(x) - y) - \psi_\tau(\pi(f_{\lambda}^{(0)})(x) - y)$ on (Z, ρ) . It satisfies $0 \leq \xi \leq |\pi(f_{\lambda}^{(0)})(x) - f_{\lambda}^{(0)}(x)| \leq 1 + \|f_{\lambda}^{(0)}\|_\infty$. Hence, $|\xi - \mathbf{E}(\xi)| \leq 1 + \|f_{\lambda}^{(0)}\|_\infty$ and $\mathbf{E}(\xi - \mathbf{E}(\xi))^2 \leq \mathbf{E}\xi^2 \leq (1 + \|f_{\lambda}^{(0)}\|_\infty)\mathbf{E}(\xi)$. Applying the one-side Bernstein inequality [12], we know that there exists a subset $Z_{1,\delta}$ of Z^m with measure at least $1 - (\delta/3)$ such that

$$\mathcal{S}_{2,1} \leq \frac{7 \left(1 + \left\| f_{\lambda}^{(0)} \right\|_\infty \right) \log(3/\delta)}{6m} + \mathcal{E}_\tau \left(f_{\lambda}^{(0)} \right) - \mathcal{E}_\tau \left(\pi \left(f_{\lambda}^{(0)} \right) \right), \quad \forall \mathbf{z} \in Z_{1,\delta}. \quad (4.5)$$

For $\mathcal{S}_{2,2}$, we apply the one-side Bernstein inequality again to the random variable $\xi(z) = \psi_\tau(\pi(f_\lambda^{(0)})(x) - y) - \psi_\tau(f_{\rho,\tau}(x) - y)$, bound the variance by Lemma 2.6 with $f = \pi(f_\lambda^{(0)})$, and find that there exists another subset $Z_{2,\delta}$ of Z^m with measure at least $1 - (\delta/3)$ such that

$$\begin{aligned} \mathcal{S}_{2,2} \leq & \frac{4 \log(3/\delta)}{3m} + \left(\frac{\theta}{2}\right)^{\theta/(4-2\theta)} \left(1 - \frac{\theta}{2}\right) \left(\frac{2C_\theta \log(3/\delta)}{m}\right)^{1/(2-\theta)} \\ & + \xi_\tau(\pi(f_\lambda^{(0)})) - \xi_\tau(f_{\rho,\tau}), \quad \forall \mathbf{z} \in Z_{2,\delta}. \end{aligned} \quad (4.6)$$

Next, we estimate the first part \mathcal{S}_1 of the sample error. Consider the function set

$$\mathcal{G} = \{\psi_\tau(\pi(f)(x) - y) - \psi_\tau(f_{\rho,\tau}(x) - y) : \|f\|_K \leq R\}. \quad (4.7)$$

A function from this set $g(z) = \psi_\tau(\pi(f)(x) - y) - \psi_\tau(f_{\rho,\tau}(x) - y)$ satisfies $\mathbf{E}(g) \geq 0$, $|g(z)| \leq 2$, and $\mathbf{E}(g^2) \leq C_\theta(\mathbf{E}(g))^\theta$ by (2.16). Also, the Lipschitz property of the pinball loss yields $\mathcal{N}(\mathcal{G}, u) \leq \mathcal{N}(B_1, u/R)$. Then, we apply a standard covering number argument with a ratio inequality [12, 13, 19, 20] to \mathcal{G} and find from the covering number condition (2.5) that

$$\begin{aligned} \text{Prob}_{\mathbf{z} \in Z^m} \left\{ \sup_{\|f\|_K \leq R} \frac{[\xi_\tau(\pi(f)) - \xi_\tau(f_{\rho,\tau})] - [\xi_{\mathbf{z},\tau}(\pi(f)) - \xi_{\mathbf{z},\tau}(f_{\rho,\tau})]}{\sqrt{(\xi_\tau(\pi(f)) - \xi_\tau(f_{\rho,\tau}))^\theta + u^\theta}} \leq 4u^{1-(\theta/2)} \right\} \\ \geq 1 - \mathcal{N}\left(B_1, \frac{u}{R}\right) \exp\left\{-\frac{mu^{2-\theta}}{2C_\theta + (4/3)u^{1-\theta}}\right\} \geq 1 - \exp\left\{C_s\left(\frac{R}{u}\right)^s - \frac{mu^{2-\theta}}{2C_\theta + (4/3)u^{1-\theta}}\right\}. \end{aligned} \quad (4.8)$$

Setting the confidence to be $1 - (\delta/3)$, we take $u^*(R, m, \delta/3)$ to be the positive solution to the equation

$$C_s\left(\frac{R}{u}\right)^s - \frac{mu^{2-\theta}}{2C_\theta + (4/3)u^{1-\theta}} = \log \frac{\delta}{3}. \quad (4.9)$$

Then, there exists a third subset $Z_{3,\delta}$ of Z^m with measure at least $1 - (\delta/3)$ such that

$$\sup_{\|f\|_K \leq R} \frac{[\xi_\tau(\pi(f)) - \xi_\tau(f_{\rho,\tau})] - [\xi_{\mathbf{z},\tau}(\pi(f)) - \xi_{\mathbf{z},\tau}(f_{\rho,\tau})]}{\sqrt{(\xi_\tau(\pi(f)) - \xi_\tau(f_{\rho,\tau}))^\theta + (u^*(R, m, \delta/3))^\theta}} \leq 4\left(u^*\left(R, m, \frac{\delta}{3}\right)\right)^{1-(\theta/2)}, \quad \forall \mathbf{z} \in Z_{3,\delta}. \quad (4.10)$$

Thus, for $z \in \mathcal{W}(R) \cap Z_{3,\delta}$, we have

$$\begin{aligned} \mathcal{S}_1 &= \left[\mathcal{E}_\tau(\pi(f_z^{(e)})) - \mathcal{E}_\tau(f_{\rho,\tau}) \right] - \left[\mathcal{E}_{z,\tau}(\pi(f_z^{(e)})) - \mathcal{E}_{z,\tau}(f_{\rho,\tau}) \right] \\ &\leq 4 \left[u^* \left(R, m, \frac{\delta}{3} \right) \right]^{1-(\theta/2)} \sqrt{\left[\mathcal{E}_\tau(\pi(f_z^{(e)})) - \mathcal{E}_\tau(f_{\rho,\tau}) \right]^\theta + \left[u^* \left(R, m, \frac{\delta}{3} \right) \right]^\theta} \\ &\leq \left(1 - \frac{\theta}{2} \right) 4^{2/(2-\theta)} u^* \left(R, m, \frac{\delta}{3} \right) + \frac{\theta}{2} \left[\mathcal{E}_\tau(\pi(f_z^{(e)})) - \mathcal{E}_\tau(f_{\rho,\tau}) \right] + 4u^* \left(R, m, \frac{\delta}{3} \right). \end{aligned} \quad (4.11)$$

Here, we have used the elementary inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and Young's inequality. Putting this bound and (4.5), (4.6) into (3.5), we know that for $z \in \mathcal{W}(R) \cap Z_{3,\delta} \cap Z_{1,\delta} \cap Z_{2,\delta}$, there holds

$$\begin{aligned} \mathcal{E}_\tau(\pi(f_z)) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \left\| f_z^{(e)} \right\|_K^2 &\leq \frac{1}{2} \left[\mathcal{E}_\tau(\pi(f_z^{(e)})) - \mathcal{E}_\tau(f_{\rho,\tau}) \right] + \epsilon + 2\mathfrak{D}_\tau(\lambda) \\ &\quad + 20u^* \left(R, m, \frac{\delta}{3} \right) + \frac{7 \left(1 + \left\| f_\lambda^{(0)} \right\|_\infty \right) \log(3/\delta)}{6m} + \frac{4 \log(3/\delta)}{3m} + \left(\frac{2C_\theta \log(3/\delta)}{m} \right)^{1/(2-\theta)}, \end{aligned} \quad (4.12)$$

which together with Proposition 3.3 implies

$$\begin{aligned} \mathcal{E}_\tau(\pi(f_z)) - \mathcal{E}_\tau(f_{\rho,\tau}) + \lambda \left\| f_z^{(e)} \right\|_K^2 &\leq 2\epsilon + 4C_0 \lambda^{r/(1-r)} + \left(5 + 2(2C_\theta)^{1/(2-\theta)} \right) \log \frac{3}{\delta} m^{-1/(2-\theta)} \\ &\quad + 40u^* \left(R, m, \frac{\delta}{3} \right) + 2\kappa \sqrt{C_0} \log \frac{3}{\delta} \frac{\lambda^{(2r-1)/(2-2r)}}{m}. \end{aligned} \quad (4.13)$$

Here, we have used the reproducing property in \mathcal{L}_K which yields [12]

$$\|f\|_\infty \leq \kappa \|f\|_K, \quad \forall f \in \mathcal{L}_K. \quad (4.14)$$

Equation (4.9) can be expressed as

$$u^{2+s-\theta} - \frac{4}{3m} \log \frac{3}{\delta} u^{1+s-\theta} - \frac{2C_\theta}{m} \log \frac{3}{\delta} u^s - \frac{4C_s R^s}{3m} u^{1-\theta} - \frac{2C_\theta C_s R^s}{m} = 0. \quad (4.15)$$

By Lemma 7.2 in [12], the positive solution $u^*(R, m, \delta/3)$ to this equation can be bounded as

$$\begin{aligned} u^*(R, m, \delta/3) &\leq \max \left\{ \frac{6}{m} \log \frac{3}{\delta}, \left(\frac{8C_\theta}{m} \log \frac{3}{\delta} \right)^{1/(2-\theta)}, \left(\frac{6C_s R^s}{m} \right)^{1/(1+s)}, \left(\frac{8C_\theta C_s R^s}{m} \right)^{1/(2+s-\theta)} \right\} \\ &\leq \left(6 + (8C_\theta)^{1/(2-\theta)} \right) \log \frac{3}{\delta} m^{-1/(2-\theta)} + \left((6C_s)^{1/(1+s)} + (8C_\theta C_s)^{1/(2+s-\theta)} \right) \\ &\quad \times m^{-1/(2+s-\theta)} R^{s/(1+s)}. \end{aligned} \quad (4.16)$$

Thus, for $\mathbf{z} \in \mathcal{W}(R) \cap Z_{3,\delta} \cap Z_{1,\delta} \cap Z_{2,\delta}$, the desired bound (4.2) holds true. Since the measure of the set $Z_{3,\delta} \cap Z_{1,\delta} \cap Z_{2,\delta}$ is at least $1 - \delta$, our conclusion is proved. \square

5. Deriving Convergence Rates by Iteration

To apply Proposition 4.1 for error analysis, we need some $R \geq 1$ for $\mathbf{z} \in \mathcal{W}(R)$. One may choose $R = \lambda^{-1/2}$ according to

$$\|f_{\mathbf{z}}^{(\epsilon)}\|_K \leq \lambda^{-1/2}, \quad \forall \mathbf{z} \in Z^m \quad (5.1)$$

which is seen by taking $f = 0$ in (1.7). This choice is too rough. Recall from Proposition 3.3 that $\|f_{\lambda}^{(0)}\|_K \leq \sqrt{C_0} \lambda^{(2r-1)/(2-2r)}$ which is a bound for the noise-free limit $f_{\lambda}^{(0)}$ of $f_{\mathbf{z}}^{(\epsilon)}$. It is much better than $\lambda^{-1/2}$. This motivates us to try similar tight bounds for $f_{\mathbf{z}}^{(\epsilon)}$. This target will be achieved in this section by applying Proposition 4.1 iteratively. The iteration technique has been used in [13, 21] to improve learning rates.

Lemma 5.1. *Assume (2.3) with $0 < r \leq 1/2$ and (2.5) with $s > 0$. Take $\lambda = m^{-\alpha}$ with $0 < \alpha \leq 1$ and $\epsilon = m^{-\beta}$ with $0 < \beta \leq \infty$. Let $0 < \eta < 1$. If ρ has a τ -quantile of p -average type q for some $p \in (0, \infty]$ and $q \in (1, \infty)$, then for any $0 < \delta < 1$, with confidence $1 - \delta$, there holds*

$$\|f_{\mathbf{z}}^{(\epsilon)}\|_K \leq 4C'_3 \left(1 + \sqrt{2} + \sqrt{C'_1} + \sqrt{C'_2}\right) \left(\log \frac{3}{\eta}\right)^2 \sqrt{\log \frac{3}{\delta}} m^{\theta_\eta}, \quad (5.2)$$

where θ_η is given by

$$\theta_\eta = \max \left\{ \frac{\alpha - \beta}{2}, \frac{\alpha(1-2r)}{2-2r}, \frac{\alpha}{2} - \frac{1}{2(2-\theta)}, \frac{[\alpha(2+s-\theta) - 1](1+s)}{(2+s-\theta)(2+s)} + \eta \right\} \geq 0. \quad (5.3)$$

Proof. Putting $\lambda = m^{-\alpha}$ with $0 < \alpha \leq 1$ and $\epsilon = m^{-\beta}$ with $0 < \beta \leq \infty$ into Proposition 4.1, we know that for any $R \geq 1$ there exists a subset V_R of Z^m with measure at most δ such that

$$\|f_{\mathbf{z}}^{(\epsilon)}\|_K \leq a_m R^{s/(2+2s)} + b_m, \quad \forall \mathbf{z} \in \mathcal{W}(R) \setminus V_R, \quad (5.4)$$

where with $\zeta := \max\{\alpha - \beta, \alpha(1-2r)/(2-2r), \alpha/2 - 1/2(2-\theta)\} \geq 0$, the constants are given by

$$a_m = \sqrt{C'_3} m^{(\alpha/2) - (1/2)(2+s-\theta)}, \quad b_m = \left(\sqrt{2} + \sqrt{C'_1 \log \frac{3}{\delta}} + \sqrt{C'_2 \log \frac{3}{\delta}} \right) m^\zeta. \quad (5.5)$$

It follows that

$$\mathcal{W}(R) \subseteq \mathcal{W}(a_m R^{s/(2+2s)} + b_m) \cup V_R. \quad (5.6)$$

Let us apply (5.6) iteratively to a sequence $\{R^{(j)}\}_{j=0}^J$ defined by $R^{(0)} = \lambda^{-1/2}$ and $R^{(j)} = a_m(R^{(j-1)})^{s/(2+2s)} + b_m$ where $J \in \mathbb{N}$ will be determined later. Then, $\mathcal{W}(R^{(j-1)}) \subseteq \mathcal{W}(R^{(j)}) \cup V_{R^{(j-1)}}$. By (5.1), $\mathcal{W}(R^{(0)}) = Z^m$. So we have

$$Z^m = \mathcal{W}(R^{(0)}) \subseteq \mathcal{W}(R^{(1)}) \cup V_{R^{(0)}} \subseteq \dots \subseteq \mathcal{W}(R^{(J)}) \cup \left(\bigcup_{j=0}^{J-1} V_{R^{(j)}} \right). \quad (5.7)$$

As the measure of $V_{R^{(j)}}$ is at most δ , we know that the measure of $\bigcup_{j=0}^{J-1} V_{R^{(j)}}$ is at most $J\delta$. Hence, $\mathcal{W}(R^{(J)})$ has measure at least $1 - J\delta$.

Denote $\Delta = s/(2+2s) < 1/2$. The definition of the sequence $\{R^{(j)}\}_{j=0}^J$ tells us that

$$R^{(J)} = a_m^{1+\Delta+\Delta^2+\dots+\Delta^{J-1}} \left(R^{(0)} \right)^{\Delta^J} + \sum_{j=1}^{J-1} a_m^{1+\Delta+\Delta^2+\dots+\Delta^{j-1}} b_m^{\Delta^j} + b_m. \quad (5.8)$$

Let us bound the two terms on the right-hand side.

The first term equals

$$\left(C'_3 \right)^{(1-\Delta^J)/2(1-\Delta)} m^{((\alpha(2+s-\theta)-1)/(4+2s-2\theta))((1-\Delta^J)/(1-\Delta))} m^{(\alpha/2)\Delta^J}, \quad (5.9)$$

which is bounded by

$$\begin{aligned} & C'_3 m^{(\alpha(2+s-\theta)-1)/(4+2s-2\theta)(1-\Delta)} m^{((\alpha/2)-(\alpha(2+s-\theta)-1)/(4+2s-2\theta)(1-\Delta))\Delta^J} \\ & \leq C'_3 m^{[\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)} m^{(1/(2+s-\theta))2^{-J}}. \end{aligned} \quad (5.10)$$

Take J to be the smallest integer greater than or equal to $\log(1/\eta)/\log 2$. The above expression can be bounded by $C'_3 m^{[\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)+\eta}$.

The second terms equals

$$\sum_{j=1}^{J-1} a_m^{1+\Delta+\Delta^2+\dots+\Delta^{j-1}} b_m^{\Delta^j} + b_m \leq \sum_{j=1}^{J-1} C'_3 m^{((\alpha/2)-1/2(2+s-\theta))((1-\Delta^j)/(1-\Delta))} b_1^{\Delta^j} m^{\zeta \Delta^j} + b_1 m^{\zeta}, \quad (5.11)$$

where $b_1 := \sqrt{2} + \sqrt{C'_1 \log(3/\delta)} + \sqrt{C'_2 \log(3/\delta)}$. It is bounded by

$$m^{[\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)} C'_3 b_1 \sum_{j=0}^{J-1} m^{(\zeta - ([\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)))(s^j/(2+2s)^j)}. \quad (5.12)$$

When $\zeta \leq [\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)$, the above expression is bounded by $C'_3 b_1 J m^{[\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)}$. When $\zeta \geq [\alpha(2+s-\theta)-1](1+s)/(2+s-\theta)(2+s)$, it is bounded by $C'_3 b_1 J m^{\zeta}$.

Based on the above discussion, we obtain

$$R^{(J)} \leq (C'_3 + C'_3 b_1 J) m^{\theta_\eta}, \quad (5.13)$$

where $\theta_\eta = \max\{[\alpha(2+s-\theta) - 1](1+s)/(2+s-\theta)(2+s) + \eta, \zeta\}$. So with confidence $1 - J\delta$, there holds

$$\|f_z^{(\varepsilon)}\|_K \leq R^{(J)} \leq C'_3 \left(1 + \sqrt{2} + \sqrt{C'_1} + \sqrt{C'_2}\right) J \sqrt{\log \frac{3}{\delta}} m^{\theta_\eta}. \quad (5.14)$$

Then, our conclusion follows by replacing δ by δ/J and noting $J \leq 2 \log(3/\eta)$. \square

Now, we can prove our main result, Theorem 2.5.

Proof of Theorem 2.5. Take R to be the right side of (5.2). By Lemma 5.1, there exists a subset V'_R of Z^m with measure at most δ such that $Z^m \setminus V'_R \subseteq \mathcal{W}(R)$. Applying Proposition 4.1 to this R , we know that there exists another subset V_R of Z^m with measure at most δ such that for any $\mathbf{z} \in \mathcal{W}(R) \setminus V_R$,

$$\begin{aligned} \mathcal{E}_\tau(\mathcal{X}(f_z^{(\varepsilon)})) - \mathcal{E}_\tau(f_{\rho,\tau}) &\leq 2m^{-\beta} + C'_1 \log \frac{3}{\delta} m^{-ar/(1-r)} + C'_2 \log \frac{3}{\delta} m^{-1/(2-\theta)} \\ &\quad + C'_4 \left(\log \frac{3}{\eta}\right)^2 \sqrt{\log \frac{3}{\delta}} m^{(s/(1+s))\theta_\eta - (1/(2+s-\theta))}, \end{aligned} \quad (5.15)$$

where

$$C'_4 = C'_3 (4C'_3)^{s/(1+s)} \left(1 + \sqrt{2} + \sqrt{C'_1} + \sqrt{C'_2}\right). \quad (5.16)$$

Since the set $V_R \cup V'_R$ has measure at most 2δ , after scaling 2δ to δ and setting the constant \tilde{C} by

$$\tilde{C} = 2C_{q,\rho} (2 + C'_1 + C'_2 + C'_4), \quad (5.17)$$

we see that the above estimate together with Lemma 3.1 gives the error bound

$$\|\mathcal{X}(f_z^{(\varepsilon)}) - f_{\rho,\tau}\|_{L_{\rho_X}^p} \leq \tilde{C} \left(\log \frac{3}{\eta}\right)^2 \log \frac{3}{\delta} m^{-\vartheta} \quad (5.18)$$

with confidence $1 - \delta$ and the power index ϑ give by

$$\vartheta = \frac{1}{q} \min\left\{\beta, \frac{ar}{1-r}, \frac{1}{2+s-\theta} - \frac{s}{1+s}\theta_\eta\right\}, \quad (5.19)$$

provided that

$$\theta_\eta < \frac{1+s}{s(2+s-\theta)}. \quad (5.20)$$

Since $\beta \geq \alpha r / (1-r)$, we know that $(\alpha - \beta) / 2 \leq \alpha(1-2r) / (2-2r)$. By the restriction $0 < \alpha < (2+s) / s(2+s-\theta)$ on α , we find $\alpha(1-2r) / (2-2r) < (1+s) / s(2+s-\theta)$ and $(\alpha/2) - (1/2(2-\theta)) < (1+s) / s(2+s-\theta)$. Moreover, restriction (2.7) on η tell us that $[\alpha(2+s-\theta) - 1](1+s) / (2+s-\theta)(2+s) + \eta < (1+s) / s(2+s-\theta)$. Therefore, condition (5.20) is satisfied. The restriction $\beta \geq \alpha r / (1-r)$ and the above expression for ϑ tells us that the power index for the error bound can be exactly expressed by formula (2.9). The proof of Theorem 2.5 is complete. \square

Finally, we prove Theorem 2.3.

Proof of Theorem 2.3. Since $f_{\rho,\tau} \in \mathcal{H}_K$, we know that (2.3) holds with $r = 1/2$. The noise condition on ρ is satisfied with $q = 2$ and $p \in (0, \infty]$. Then, $\theta = p / (p+1) \in (0, 1]$. Since $X \subset \mathbb{R}^n$ and $K \in C^\infty(X \times X)$, we know from [11] that (2.5) holds true for any $s > 0$. With $0 < \eta < ((p+1)/2(p+2))$, let us choose s to be a positive number satisfying the following four inequalities:

$$\begin{aligned} \frac{p+1}{p+2} &< \frac{2+s}{s(2+s-\theta)}, \\ \frac{1}{3} &< \frac{(1+s)[2+2s-(s(2+s-\theta)(p+1))]/(p+2)}{s(2+s-\theta)(2+s)}, \\ \frac{p+1}{p+2} - 2\eta &\leq \frac{1}{2+s-\theta} - \frac{s[(2+s-\theta)(p+1)/(p+2-1)]}{(2+s-\theta)(2+s)} - \frac{s}{3(1+s)}, \\ \frac{p+1}{p+2} - 2\eta &\leq \frac{1}{2+s-\theta}. \end{aligned} \quad (5.21)$$

The first inequality above tells us that the restrictions on α, β are satisfied by choosing $\alpha = (p+1)/(p+2) = 1/(2-\theta)$ and $(p+1)/(p+2) \leq \beta \leq \infty$. The second inequality shows that condition (2.7) for the parameter η renamed now as η^* is also satisfied by taking $\eta^* = 1/3$. Thus, we apply Theorem 2.5 and know that with confidence $1-\delta$, (2.8) holds with the power index ϑ given by (2.9); but $r = 1/2, \alpha = 1/(2-\theta)$, and $\eta^* = 1/3$ imply that

$$\vartheta = \frac{1}{2} \min \left\{ \alpha, \frac{1}{2+s-\theta}, \frac{1}{2+s-\theta} - \frac{s[\alpha(2+s-\theta) - 1]}{(2+s-\theta)(2+s)} - \frac{s}{3(1+s)} \right\}. \quad (5.22)$$

The last two inequalities satisfied by s yield $\vartheta \geq (\alpha/2) - \eta$. So (2.8) verifies (2.4). This completes the proof of the theorem. \square

Acknowledgment

The work described in this paper is supported by National Science Foundation of China under Grant 11001247 and by a grant from the Research Grants Council of Hong Kong (Project no. CityU 103709).

References

- [1] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, NY, USA, 1998.
- [2] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.
- [3] H. Tong, D.-R. Chen, and L. Peng, "Analysis of support vector machines regression," *Foundations of Computational Mathematics*, vol. 9, no. 2, pp. 243–257, 2009.
- [4] R. Koenker, *Quantile Regression*, vol. 38 of *Econometric Society Monographs*, Cambridge University Press, Cambridge, UK, 2005.
- [5] I. Steinwart and A. Christman, "How SVMs can estimate quantile and the median," in *Advances in Neural Information Processing Systems*, vol. 20, pp. 305–312, MIT Press, Cambridge, Mass, USA, 2008.
- [6] I. Steinwart and A. Christmann, "Estimating conditional quantiles with the help of the pinball loss," *Bernoulli*, vol. 17, no. 1, pp. 211–225, 2011.
- [7] D. H. Xiang, "Conditional quantiles with varying Gaussians," online first in *Advances in Computational Mathematics*.
- [8] T. Hu, D. H. Xiang, and D. X. Zhou, "Online learning for quantile regression and support vector regression," Submitted.
- [9] D.-R. Chen, Q. Wu, Y. Ying, and D.-X. Zhou, "Support vector machine soft margin classifiers: error analysis," *Journal of Machine Learning Research*, vol. 5, pp. 1143–1175, 2004.
- [10] D.-X. Zhou, "The covering number in learning theory," *Journal of Complexity*, vol. 18, no. 3, pp. 739–767, 2002.
- [11] D.-X. Zhou, "Capacity of reproducing kernel spaces in learning theory," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1743–1752, 2003.
- [12] F. Cucker and D.-X. Zhou, *Learning Theory: An Approximation Theory Viewpoint*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2007.
- [13] 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.
- [14] Q. Wu and D.-X. Zhou, "Analysis of support vector machine classification," *Journal of Computational Analysis and Applications*, vol. 8, no. 2, pp. 99–119, 2006.
- [15] T. Hu, "Online regression with varying Gaussians and non-identical distributions," *Analysis and Applications*, vol. 9, no. 4, pp. 395–408, 2011.
- [16] D.-H. Xiang, T. Hu, and D.-X. Zhou, "Learning with varying insensitive loss," *Applied Mathematics Letters*, vol. 24, no. 12, pp. 2107–2109, 2011.
- [17] 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.
- [18] S. Smale and D.-X. Zhou, "Online learning with Markov sampling," *Analysis and Applications*, vol. 7, no. 1, pp. 87–113, 2009.
- [19] Y. Yao, "On complexity issues of online learning algorithms," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6470–6481, 2010.
- [20] Y. Ying, "Convergence analysis of online algorithms," *Advances in Computational Mathematics*, vol. 27, no. 3, pp. 273–291, 2007.
- [21] I. Steinwart and C. Scovel, "Fast rates for support vector machines using Gaussian kernels," *The Annals of Statistics*, vol. 35, no. 2, pp. 575–607, 2007.