Open Journal of Statistics, 2017, 7, 786-802
http://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

# Quantile Regression Based on Laplacian Manifold Regularizer with the Data Sparsity in $l_1$ Spaces

**Ru Feng[1], Shuang Chen[2*], Lanlan Rong[1]**

[1]School of Sciences, Hebei University of Technology, Tianjin, China
[2]School of Sciences, Beijing Institute of Petrochemical Technology, Beijing, China
Email: *shchen@bipt.edu.cn

**How to cite this paper:** Feng, R., Chen, S. and Rong, L.L. (2017) Quantile Regression Based on Laplacian Manifold Regularizer with the Data Sparsity in $l_1$ Spaces. *Open Journal of Statistics*, **7**, 786-802.
https://doi.org/10.4236/ojs.2017.75056

**Received:** August 30, 2017
**Accepted:** October 20, 2017
**Published:** October 23, 2017



## Abstract

In this paper, we consider the regularized learning schemes based on $l_1$-regularizer and pinball loss in a data dependent hypothesis space. The target is the error analysis for the quantile regression learning. There is no regularized condition with the kernel function, excepting continuity and boundness. The graph-based semi-supervised algorithm leads to an extra error term called manifold error. Part of new error bounds and convergence rates are exactly derived with the techniques consisting of $l_1$-empirical covering number and boundness decomposition.

## Keywords

Semi-Supervised Learning, Conditional Quantile Regression, $l_1$-Regularizer, Manifold-Regularizer, Pinball Loss

## 1. Introduction

The classical least-squares regression models have focused mainly on estimating conditional mean functions. In contrast, quantile regression can provide richer information about the conditional distribution of response variables such as stretching or compressing tails, so it is particularly useful in applications when both lower and upper or all quantiles are of interest. Over the last years, quantile regression has become a popular statistical method in various research fields, such as reference charts in medicine [1], survival analysis [2] and economics [3].

In addition, relative to the least-squares regression, quantile regression estimates are more robust against outliers in the response measurements. We introduce a framework for data-dependent regularization that exploits the geometry of the marginal distribution. The labeled and unlabeled data learnt from the problem constructs a framework and incorporates the framework as an additional regularization term. The framework exploits the geometry of the probability distribution that generates the data. Hence, there are two regularization terms: one controlling the complexity of the classifier in the ambient space and the other controlling the complexity as measured by geometry of the distribution in the intrinsic space.

## 2. The Model

In this paper, under the framework of learning theory, we study $l_1$-regularized and manifold regularized quantile regression. Let $X$ be a compact subset of $\mathbb{R}^{l+u}$ and $Y \subset \mathbb{R}$. There is a probability distribution $\rho$ on $X \times Y$ according to which examples are generated for function learning. Labeled examples are $(x, y)$ pairs generated according to $\rho$. Unlabeled examples are simply $x \in X$ drawn according to the marginal distribution $\rho_X$ of $\rho$. We will make a specific assumption that an identifiable relation between $\rho_X$ and the conditional $\rho(y \mid x)$. The conditional $\tau$-quantile is a set-valued function defined by

$$\rho\left(\left\{y \in (-\infty, u]\right\} \mid x\right) \geq 1-\tau \text{ and } \rho\left(\left\{y \in (u, \infty]\right\} \mid x\right) \geq \tau \tag{2.1}$$

where $\tau \in (0,1)$, $x \in X$ and $u \in Y$.

The empirical method for estimating the conditional $\tau$-quantile function is based on the $\tau$-pinball loss

$$\rho_\tau(r) = \begin{cases} \tau r, & r > 0 \\ (\tau - 1) r, & \text{otherwise} \end{cases} \tag{2.2}$$

Then, denote the generalization error to minimize the conditional $\tau$-quantile function $f_\rho$ with the loss function $\rho_\tau$

$$\varepsilon^\tau = \int_{X \times Y} \rho_\tau\left(y - f(x)\right) \mathrm{d}\rho \tag{2.3}$$

where $f : X \to \mathbb{R}$. Based on observations, the empirical risk of the function $f$ is

$$\varepsilon_z^\tau = \frac{1}{l} \sum_{i=1}^{l} \rho_\tau\left(y_i - f(x_i)\right) \tag{2.4}$$

Next, We assume that $\left|f_\rho^\tau\right| \leq 1, a.e., x \in X$ with respect to $\rho_X$.

In kernel-based learning, this minimization process usually takes place in a hypothesis space, Reproducing Kernel Hilbert Space (RKHS) [4] [5] $\mathcal{H}_k$ generated by a kernel function $K : X \times X \to \mathbb{R}$. In the empirical case, a graph-based regular quantile regression problem can be typically formulated in terms of the following optimization

$$\hat{f}_{z,\gamma} = \min_{f \in \mathcal{H}_K} \left\{ \sum_{i=1}^{l} \rho_\tau\left(y_i - f(x_i)\right) + \gamma_A \|f\|_K + \gamma_I \|f\|_I \right\}. \tag{2.5}$$

By the representers theorem, the solution to (2.5) can be written as

$$\hat{f}_{z,\gamma}(x)=\sum_{i=1}^{l+u}\alpha_i K(x,x_i) \tag{2.6}$$

The $l_1$-norm penalty not only shrinks the fitted coefficients toward zero but also causes some of the coefficients to be exactly zero when making $\gamma_A$ sufficiently large. When the data lies on a low-dimensional manifold, the graph-based method seems more effective for semi-supervised learning and many approaches have been proposed for instance Transductive SVM [6], Measure-based Regularization [7] and so on. Then the $l_1$-regularized and manifold regularized quantile regression are as following

$$\begin{aligned}\hat{f}_{z,\gamma}(x)&=\min_{f\in\mathcal{H}_K}\left\{\varepsilon_z^\tau(f)+\gamma_A\Omega(f)+\frac{\gamma_I}{(l+u)^2}\sum_{i,j=1}^{l+u}\left(f(x_i)-f(x_j)\right)^2\omega_{ij}\right\}\\&=\min_{f\in\mathcal{H}_K}\left\{\varepsilon_z^\tau(f)+\gamma_A\Omega(f)+\frac{\gamma_I}{(l+u)^2}\mathbf{f}^{\mathrm{T}}L\mathbf{f}\right\}\end{aligned} \tag{2.7}$$

where $\Omega(f)=\sum_{i=1}^{l+u}|\alpha_i|$, $\mathbf{f}=\left(f(x_1),\cdots,f(x_{l+u})\right)^{\mathrm{T}}$. $\gamma_A,\gamma_I$ are nonnegative regularization parameters. $L=D-W$ is the unnormalized graph Laplacian, where $D$ is a diagonal matrix with diagonal entries $D_{ii}=\sum_{j=1}^{l+u}\omega_{ij}$. The weight $\omega_{ij}$ is given by a similar function $\omega(x_i,x_j)$. The more similar $x_i$ and $x_j$, the larger $\omega_{ij}$ should be.

## 3. The Restriction

**Definition 3.1.** *The projection operator on the space of function is defined by*

$$\pi(f)(x)=\begin{cases}1, & \text{if } f(x)>1\\ f(x), & \text{if } -1\le f(x)\le 1\\ -1, & \text{if } f(x)<1\end{cases} \tag{3.1}$$

Hence, it is natural to measure the approximation ability by the distance $\left\|\pi\left(\hat{f}_{z,\gamma}\right)-f_\rho^\tau\right\|_{L^{p^*}_{\rho_X}}$ with $f_\rho^\tau\le 1$.

**Definition 3.2.** *Let* $p\in(0,\infty]$ *and* $q\in(1,\infty)$. *We say that* $\rho$ *has a* $\tau$*-quantile of* $p$*-average type* $q$ *if for almost all* $x\in X$ *with respect to* $\rho_X$, *there exist a* $\tau$ *quantile* $t\in\mathbb{R}$ *and constants* $a_x\in(0,2]$, $b_x>0$ *such that for each* $s\in[0,a_x]$,

$$\rho\left(y\in(t-s,t)\,|\,x\right)\ge b_x s^{q-1}, \tag{3.2}$$

$$\rho\left(y\in(t-s,t)\,|\,x\right)\ge b_x s^{q-1}, \tag{3.3}$$

and that the function $\phi:X\to[0,\infty]$, $\phi(x)=b_x a_x^{q-1}$ satisfies $\phi_{-1}\in L^p_{\rho_X}$.

For $p\in(0,\infty]$ and $q\in(1,\infty)$, denote

$$\theta=\min\left\{\frac{2}{q},\frac{p}{p+1}\right\}\in(0,1] \tag{3.4}$$

**Lemma 3.1.** *If* $\rho$ *has a* $\tau$*-quantile of* $p$*-average type* $q$ *for some* $p\in(0,\infty]$ *and* $q\in(1,\infty)$, *then for any measurable function* $f:X\to[-1,1]$, *there holds*

$$\left\| f - f_\rho^\tau \right\|_{L^{p_*}_{\rho_X}} \le C_{q,\rho} \left\{ \varepsilon^\tau(f) - \varepsilon^\tau\left(f_\rho^\tau\right) \right\}^{\frac{1}{q}} \tag{3.5}$$

where $C_{q,\rho} = 2^{1-\frac{1}{q}} q^{\frac{1}{q}} \left\| \left[ \left( b_x a_x^{q-1} \right)^{-1} \right]_{x \in X} \right\|_{L^p_{\rho_X}}^{\frac{1}{q}}$ and $p_* = \dfrac{pq}{p+1}$

**Definition 3.3.** *We say that the kernel function* $K$ *is a* $C^c$ *kernel with* $c > 0$ *if there exists some constants* $C^c > 0$, *such that*

$$\left| K(t,x) - K(t,x') \right| \le C^c \left| x - x' \right|^c, \forall t, x, x' \in X \tag{3.6}$$

We assume throughout the paper that $K \in C^c(X \times X)$ and denote $\kappa = \sup_{t,x \in X} \left| K(x,t) \right| < \infty$. Our approximation condition is given as

$$f_\rho^\tau = L_{\tilde{K}}^s g_\rho^\tau, \ \text{for some } 0 < s < 1, \ g_\rho^\tau \in L^2_{\rho_X}(X) \tag{3.7}$$

here, the kernel $\tilde{K}$ is defined by

$$\tilde{K}(x,y) = \int_X K(x,t) K(y,t) \mathrm{d}\rho_X(t) \tag{3.8}$$

Hence, although kernel $K$ in not positive semi-definite, $\tilde{K}$ is a Mercer kernel, $\mathcal{H}_{\tilde{K}}$ denotes the associated reproducing kernel Hilbert spaces. The kernel $\tilde{K}$ defines an integral operator $L_{\tilde{K}} : L^2_{\rho_X} \to L^2_{\rho_X}$ by

$$L_{\tilde{K}} f(x) = \int_X \tilde{K}(x,x') f(x') \mathrm{d}\rho_X(x'), \ x \in X. \tag{3.9}$$

Note that $L_{\tilde{K}} = L_K L_K^*$ is a self-adjoint positive operator on $L^2_{\rho_X}$ .Hence its *s*-th power $L_{\tilde{K}}^s$ is well defined for any $s > 0$. we take the RKHS $\mathcal{H}_{\hat{K}}$ with

$$\hat{K}(X,Y) = \int_X \tilde{K}(x,t) \tilde{K}(y,t) \mathrm{d}\rho_X(t) \tag{3.10}$$

It is easy to see $L_{\hat{K}} = L_{\tilde{K}}^2$, so that any function $f \in \mathcal{H}_{\hat{K}}$ can be expressed as $L_{\tilde{K}} g$ for some $g \in L^2_{\rho_X}$.

**Definition 3.4.** *Define a Banach space* $\mathcal{H}_1 = \left\{ f : f = \sum_{j=1}^\infty \alpha_j \hat{K}(x, x_j), \{\alpha_j\} \in l_1, \{x_j\} \subset X \right\}$ *with the norm*

$$\|f\| = \inf \left\{ f = \sum_{j=1}^\infty \left| \alpha_j \right| : f = \sum_{j=1}^\infty \alpha_j \hat{K}(x, x_j), \{\alpha_j\} \in l_1, \{x_j\} \subset X \right\} \tag{3.11}$$

**Definition 3.5.** *For every* $\eta > 0$, *the* $l_2$*-empirical covering number of* $\mathcal{F}$ *is*

$$\begin{aligned} \mathcal{N}_2(\mathcal{F}, \eta) = \min_{k \in \mathbb{N}} \min_{x \in X^k} \inf \Big\{ & l \in \mathbb{N} : \exists \{f_i\}_{i=1}^l \text{ such that for all} \\ & f \in \mathcal{F}, \text{there is} \min_{1 \le i \le l} d_{2,x}(f, f_i) \le \eta \Big\} \end{aligned} \tag{3.12}$$

**Lemma 3.2.** *There exist an exponent* $\mu \in (0,2)$ *and a constant* $c_{\mu,K}$ *such that*

$$\log \mathcal{N}_2(B_r, \eta) \le c_{\mu,K} \eta^{-\mu}, \forall \eta > 0 \tag{3.13}$$

suppose that $K \in C^c(X \times X)$

$$\mu=\begin{cases}2n/(n+2c), & \text{when } 0<c<1\\ 2n/(n+2), & \text{when } 0<c<1\\ n/c, & \text{when } c>1+n/2\end{cases} \quad (3.14)$$

Define an operator $L_\omega$ on $L^2_{\rho_X}$ as

$$L_\omega f(x)=f(x)p(x)-\int_X K(x,x')f(x')\mathrm{d}\rho_X(x') \quad (3.15)$$

with $p(x)=\int_X K(x,x')\mathrm{d}\rho(x')$. The above equation tells us that

$$\langle f,L_\omega f\rangle_2=\iint\left(f(x)-f(x')\right)^2\mathrm{d}\rho_X(x)\mathrm{d}\rho_X(x') \quad (3.16)$$

Next, The performance of $\mathcal{H}_{\hat{K}}$ approaching $f_\rho^\tau$ can be described through the regularizing function $f_\gamma$ defined as

$$f_\gamma=\arg\min_{f\in\mathcal{H}_{\hat{K}}}\left\{\varepsilon^\tau(f)-\varepsilon(f_\rho^\tau)+\gamma_A\|f\|_{\hat{K}}+\gamma_I\langle f,L_\omega f\rangle\right\} \quad (3.17)$$

the above function $f_\gamma$ given by (3.13 ) can be expressed as

$$f_\gamma=L_{\tilde{K}}h_\gamma=L_K g_\gamma \quad (3.18)$$

where $g_\gamma=L_K^* h_\gamma$. Moreover, $g_\gamma$ is a continuous function on $X$ and

$$\|g_\gamma\|_{L^2_{\rho_X}}=\|f_\gamma\|_{\tilde{K}},\|f_\gamma\|_{\tilde{K}}\le\kappa\|f_\gamma\|_{\hat{K}}=\kappa\|h_\gamma\|_{L^2_{\rho_X}}. \quad (3.19)$$

**Definition 3.6.** *Let* $\mathcal{F}$ *be a set of function on* $X$, $\mathrm{x}=(x_i)_{i=1}^k\in X^k$. *The* $l_2$ *metric between function on* $X$ *is*

$$d_{2,\mathrm{x}}(f,g)=\left\{\frac{1}{k}\sum_{i=1}^k\left(f(x_i)-g(x_i)\right)^2\right\}^{\frac{1}{2}},\forall f,g\in\mathcal{F} \quad (3.20)$$

Denote the ball of radius $r\ge1$ as $B_r=\left\{f\in\mathcal{H}_{\hat{K}}:\|f\|_{\hat{K}}\le r\right\}$.

# 4. Error Analysis

## 4.1. Error Decomposition

**Proposition 4.1.** *Let* $\gamma=(\gamma_A,\gamma_I),\gamma_A>0,\gamma_I>0$ *and* $f_{z,\gamma}=\sum_{i=1}^{l+u}\alpha_i K(x,x_i)$ *given by* (2.6). *Then*

$$\begin{aligned}&\varepsilon^\tau\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon^\tau\left(f_\rho^\tau\right)+\gamma_A\Omega\left(f_{z,\gamma}\right)+\frac{\gamma_I}{(l+u)^2}f_{z,\gamma}^{\mathrm{T}}Lf_{z,\gamma}\\&\le\mathcal{S}_1(z,\gamma)+\mathcal{S}_2(z,\gamma)+\mathcal{H}_1(z,\gamma)+\mathcal{H}_2(z,\gamma)+(1+\kappa)\mathcal{D}(\gamma)\end{aligned} \quad (4.1.1)$$

where

$$\mathcal{S}_1(z,\gamma)=\left\{\varepsilon^\tau\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\}-\left\{\varepsilon_z^\tau\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon_z^\tau\left(f_\rho^\tau\right)\right\} \quad (4.1.2)$$

$$\mathcal{S}_2(z,\gamma)=\left\{\varepsilon_z^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon_z^\tau\left(f_\rho^\tau\right)\right\}-\left\{\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\} \quad (4.1.3)$$

$$\mathcal{H}_1(z,\gamma)=\gamma_A\Omega\left(\hat{f}_{z,\gamma}\right)+\frac{\gamma_I}{(l+u)^2}\hat{f}_{z,\gamma}^{\mathrm{T}}L\hat{f}_{z,\gamma}-\gamma_A\|g_\gamma\|_{L^1_{\rho_X}}-\frac{\gamma_I}{(l+u)^2}f_\gamma^{\mathrm{T}}Lf_\gamma \quad (4.1.4)$$

$$\mathcal{H}_2(z,\gamma)=\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\gamma\right) \quad (4.1.5)$$

$$\mathcal{D}(\gamma)=\varepsilon^{\tau}\left(f_{\gamma}\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)+\gamma_{A}\left\|f_{\gamma}\right\|_{K}+\gamma_{I}\left\langle f_{\gamma},L_{\omega}f_{\gamma}\right\rangle_{2} \tag{4.1.6}$$

$$\mathcal{M}(z,\gamma)=\frac{\gamma_{I}}{(l+u)^{2}}f_{\gamma}^{\mathrm{T}}Lf_{\gamma}-\gamma_{I}\left\langle f_{\gamma},L_{\omega}f_{\gamma}\right\rangle_{2} \tag{4.1.7}$$

*Proof.* A direct decomposition shows that

$$\begin{aligned}
&\varepsilon^{\tau}\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)+\gamma_{A}\Omega\left(f_{z,\gamma}\right)+\frac{\gamma_{I}}{(l+u)^{2}}f_{z,\gamma}^{\mathrm{T}}Lf_{z,\gamma}\\
&=\left\{\varepsilon^{\tau}\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)\right\}-\left\{\varepsilon_{z}^{\tau}\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon_{z}^{\tau}\left(f_{\rho}^{\tau}\right)\right\}\\
&\quad+\left\{\varepsilon^{\tau}\left(\pi\left(f_{z,\gamma}\right)\right)+\gamma_{A}\Omega\left(f_{z,\gamma}\right)+\frac{\gamma_{I}}{(l+u)^{2}}f_{z,\gamma}^{\mathrm{T}}Lf_{z,\gamma}\right\}\\
&\quad-\left\{\varepsilon_{z}^{\tau}\left(\hat{f}_{z,\gamma}\right)+\gamma_{A}\Omega\left(\hat{f}_{z,\gamma}\right)+\frac{\gamma_{I}}{(l+u)^{2}}\hat{f}_{z,\gamma}^{\mathrm{T}}L\hat{f}_{z,\gamma}\right\}\\
&\quad+\left\{\varepsilon_{z}^{\tau}\left(\hat{f}_{z,\gamma}\right)-\varepsilon_{z}^{\tau}\left(f_{\rho}^{\tau}\right)\right\}-\left\{\varepsilon^{\tau}\left(\hat{f}_{z,\gamma}\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)\right\}+\varepsilon^{\tau}\left(\hat{f}_{z,\gamma}\right)\\
&\quad-\varepsilon^{\tau}\left(f_{\gamma}\right)+\varepsilon^{\tau}\left(f_{\gamma}\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)+\gamma_{A}\left\|f_{\gamma}\right\|_{K}+\gamma_{I}\left\langle f_{\gamma},L_{\omega}f_{\gamma}\right\rangle_{2}\\
&\quad+\gamma_{A}\Omega\left(\hat{f}_{z,\gamma}\right)+\frac{\gamma_{I}}{(l+u)^{2}}\hat{f}_{z,\gamma}^{\mathrm{T}}L\hat{f}_{z,\gamma}-\gamma_{A}\left\|g_{\gamma}\right\|_{L_{\rho_X}^{1}}-\frac{\gamma_{I}}{(l+u)^{2}}f_{\gamma}^{\mathrm{T}}Lf_{\gamma}\\
&\quad+\gamma_{A}\left\|g_{\gamma}\right\|_{L_{\rho_X}^{1}}-\gamma_{A}\left\|g_{\gamma}\right\|_{L_{\rho_X}^{2}}+\frac{\gamma_{I}}{(l+u)^{2}}f_{\gamma}^{\mathrm{T}}Lf_{\gamma}-\gamma_{I}\left\langle f_{\gamma},L_{\omega}f_{\gamma}\right\rangle_{2}
\end{aligned} \tag{4.1.8}$$

The fact $|y|<1$ implies that $\varepsilon_{z}^{\tau}\left(\pi\left(f_{z,\gamma}\right)\right)\le\varepsilon_{z}^{\tau}\left(f_{z,\gamma}\right)$. Hence, the second item in the right-hand side of the above equation is at most 0 by the reason that $\hat{f}_{z,\gamma}\in\mathcal{H}_{K}$ is the minimizer of (2.7). Duo to $\left\|g_{\gamma}\right\|_{L_{\rho_X}^{1}}\le\left\|g_{\gamma}\right\|_{L_{\rho_X}^{2}}$, we see that the last but one item is at most 0. The fifth item is less than by the $(1+\kappa)\mathcal{D}(\gamma)$ fact that $\left\|g_{\gamma}\right\|_{L_{\rho_X}^{2}}=\left\|f_{\gamma}\right\|_{\tilde{K}}\le\kappa\left\|f_{\gamma}\right\|_{K}$. Thus we complete the proof.□

## 4.2. Estimation of the Regularization Error

**Proposition 4.2.** *Assume* (3.7) *holds, denoting* $\gamma_{I}=\gamma_{A}^{2-s}$ *for some* $0<s\le1$ *then we have*

$$\mathcal{D}(\gamma)\le C_{0}\gamma_{A}^{s} \tag{4.2.1}$$

where $C_{0}=2\left\|g_{\rho}^{\tau}\right\|_{L_{\rho_X}^{2}}+6\omega\left\|g_{\rho}^{\tau}\right\|_{L_{\rho_X}^{2}}^{2}$.

*Proof.* Denote $f_{\gamma_A}=\arg\min_{f\in\mathcal{H}_K}\left\{\varepsilon^{\tau}(f)+\gamma_{A}\|f\|_{K}\right\}$. By proposition 2 in [8], we get the following relationships $\varepsilon^{\tau}\left(f_{\gamma_A}\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)+\gamma_{A}\left\|f_{\gamma_A}\right\|_{K}\le2\left\|g_{\rho}^{\tau}\right\|_{L_{\rho_X}^{2}}\gamma_{A}^{s}$ and $\left\|f_{\gamma_A}\right\|_{K}\le\gamma_{A}^{s-1}\left\|g_{\rho}^{\tau}\right\|_{L_{\rho_X}^{2}}$. Connected with the definition of $\mathcal{D}(\gamma)$, we have

$$\begin{aligned}\mathcal{D}(\gamma) &\le \varepsilon^{\tau}\left(f_{\gamma_A}\right)-\varepsilon^{\tau}\left(f_{\rho}^{\tau}\right)+\gamma_A\left\|f_{\gamma}\right\|_K+\gamma_I\left\langle f_{\gamma},L_{\omega}f_{\gamma}\right\rangle_2\\ &\le 2\left\|g_{\rho}^{\tau}\right\|_{L^2_{\rho_X}}\ \gamma_A^s+\gamma_I\iint\left(f_{\gamma_A}(x)-f_{\gamma_A}(x')\right)^2\omega(x,x')\mathrm{d}\rho_X(x)\mathrm{d}\rho_X(x')\\ &\le 2\left\|g_{\rho}^{\tau}\right\|_{L^2_{\rho_X}}\ \gamma_A^s+\gamma_A^{2-s}\iint 2\left\|f_{\gamma_A}\right\|_K^2\omega\mathrm{d}\rho_X(x)\mathrm{d}\rho_X(x')\\ &\le 2\left\|g_{\rho}^{\tau}\right\|_{L^2_{\rho_X}}\ \gamma_A^s+6\omega\gamma_A^{2(s-1)}\left\|g_{\rho}^{\tau}\right\|_{L^2_{\rho_X}}\end{aligned}$$

where $\gamma_I=\gamma_A^{2-s}$. we derive the desired bound.□

### 4.3. Estimation of the Manifold Error

In this subsection, we estimation the manifold error. Denote

$$A_{z,\gamma}=\frac{\gamma_I}{l+u}\sum_{j=1}^{l+u}\left(\left(\frac{1}{l+u}\sum_{i=1}^{l+u}\xi_1(x_i)\right)(x_j)-(E\xi_1)(x_j)\right) \tag{4.3.1}$$

$$B_{z,\gamma}=\gamma_I\int f_{\gamma}^2(x)\left(\left(\frac{1}{l+u}\sum_{i=1}^{l+u}\xi_2(x_j)\right)(x)-(E\xi_2)(x)\right)\mathrm{d}\rho_X(x) \tag{4.3.2}$$

$$C_{z,\gamma}=\frac{\gamma_I}{l+u}\sum_{j=1}^{l+u}f_{\gamma}(x_j)\left((E\xi_3)(x_j)-\left(\frac{1}{l+u}\sum_{i=1}^{l+u}\xi_3(x_i)(x_j)\right)\right) \tag{4.3.3}$$

$$D_{z,\gamma}=\gamma_I\int f_{\gamma}(x)\left((E\xi_3)(x)\left(\frac{1}{l+u}\sum_{i=1}^{l+u}\xi_3(x_j)\right)(x)\right)\mathrm{d}\rho_X(x) \tag{4.3.4}$$

where $\xi_1(x)=f_{\gamma}^2(x)\omega(x,\cdot)$, $\xi_2(x)=\omega(\cdot,x)$, $\xi_3(x)=f_{\gamma}(x)\omega(\cdot,x)$. So we can see that $\mathcal{M}(z,\gamma)=2\left(A_{z,\gamma}+B_{z,\gamma}+C_{z,\gamma}+D_{z,\gamma}\right)$

**Lemma 4.1.** *Let* $\xi$ *be a random variable on a probability space* $X$ *with* $\sigma^2=E\|\xi\|^2$ *satisfying* $\|\xi\|\le M_{\xi}$ *for some constant* $M_{\xi}$. *Then for any* $0<\delta<1$, *we have*

$$\frac{1}{l}\sum_{i=1}^{l}\xi(z_i)-E\xi\le\frac{2M_{\xi}\log(1/\delta)}{l}+\sqrt{\frac{2\sigma^2\log(1/\delta)}{l}} \tag{4.3.5}$$

**Proposition 4.3.** *under the approximation condition* (3.13), *let* $0<\gamma_A\le 1$ *and* $\gamma_I=\gamma_A^{2-s}$ *for some* $0<s\le 1$. *then for any* $0<\delta<1$ *with the confidence* $1-\delta$, *there holds*

$$\mathcal{M}(z,\gamma)\le 4\omega\kappa^4C_0^2\sqrt{2\log(4/\delta)}\gamma_A^s(l+u)^{-1/2} \tag{4.3.6}$$

*Proof.* By the definition of $\xi_1(x)$, we have $|\xi_1(x)|\le\omega|f|_{\gamma}^2=\omega|L_K g_{\gamma}|^2\le\kappa\|g_{\gamma}\|_{L^2_{\rho_X}}^2$. Since $\|g_{\gamma}\|_{L^2_{\rho_X}}^2\le\kappa\|f_{\gamma}\|_{\hat{K}}\le\kappa C_0\gamma_A^{s-1}$, there holds $|\xi_1(x)|\le\kappa^4C_0^2\gamma_A^{2(s-1)}$, $|\xi_2(x)|\le\omega$, $|\xi_3(x)|\le\omega\kappa^2C_0\gamma_A^{s-1}$. Applying lemma 4.1, with confidence $1-4/\delta$,

$$A_{z,\gamma}\le\gamma_I\omega\kappa^4C_0^2\gamma_A^{2(s-1)}\left(\frac{2\log(2/\delta)}{l+u}+\sqrt{\frac{2\log(2/\delta)}{l+u}}\right) \tag{4.3.7}$$

$$B_{z,\gamma}\le\gamma_I\omega\kappa^4C_0^2\gamma_A^{2(s-1)}\left(\frac{2\log(2/\delta)}{l+u}+\sqrt{\frac{2\log(2/\delta)}{l+u}}\right) \tag{4.3.8}$$

$$C_{z,\gamma} \le \gamma_I \omega \kappa^4 C_0^2 \gamma_A^{2(s-1)} \left( -\frac{2\log(2/\delta)}{l+u} + \sqrt{\frac{2\log(2/\delta)}{l+u}} \right) \tag{4.3.9}$$

$$D_{z,\gamma} \le \gamma_I \omega \kappa^4 C_0^2 \gamma_A^{2(s-1)} \left( -\frac{2\log(2/\delta)}{l+u} + \sqrt{\frac{2\log(2/\delta)}{l+u}} \right) \tag{4.3.10}$$

Then we find the manifold error bound holds true.□

## 4.4. Estimation of the Hypothesis Error

This subsection is devoted to estimate the hypothesis errors. Under the assumption that the sample is i.i.d. drawn from $\rho$ and $|y| \le 1$ *a.e.*, we estimate $\mathcal{H}_1, \mathcal{H}_2$ as following.

**Proposition 4.4.** *For any* $0 < \delta < 1$, *with confidence* $1-\delta$, *we have*

$$\mathcal{H}_1 \le \kappa \frac{\mathcal{D}(\gamma)}{\sqrt{l+u}} \left\{ \frac{2\log(4/\delta)}{\sqrt{l+u}} + \sqrt{2\log(4/\delta)} \right\} + 8\kappa^4 \sqrt{2\log(4/\delta)} \frac{\mathcal{D}(\gamma)^2}{\gamma_A^s \sqrt{l+u}} \tag{4.4.1}$$

$$\mathcal{H}_2 \le \kappa^2 \frac{\mathcal{D}(\gamma)}{\gamma_A \sqrt{l+u}} \left\{ \frac{2\log(4/\delta)}{\sqrt{l+u}} + \sqrt{2\log(4/\delta)} \right\} \tag{4.4.2}$$

*Proof.* We estimate $\mathcal{H}_1$. Recall $\hat{f}_{z,\gamma} = \frac{1}{l+u} \sum_{i=1}^{l+u} g_\gamma(x_i) K_{x_i}$, then

$$\Omega\left(\hat{f}_{z,\gamma}\right) = \frac{1}{l+u} \sum_{i=1}^{l+u} \left| g_\gamma(x_i) \right|.$$

Applying Lemma 4.1 to the random variable $\xi = |g_\gamma(x)|$ on $(X, \rho_X)$ with value in $\mathbb{R}$. There is

$$\xi = |g_\gamma(x)| \le \kappa \mathcal{D}(\gamma)/\gamma_A \tag{4.4.3}$$

$$E\xi = \|g_\gamma\|_{L^1_{\rho_X}} \text{ and } \sigma^2(\xi) \le \kappa^2 \mathcal{D}(\gamma)^2 / \gamma_A^2 \tag{4.4.4}$$

with confidence $1-\delta/2$, there holds

$$\Omega\left(\hat{f}_{z,\gamma}\right) - \|g_\gamma\|_{L^1_{\rho_X}} \le \frac{2\kappa \mathcal{D}(\gamma) \log(4/\delta)}{\gamma_A (l+u)} + \sqrt{\frac{2\kappa^2 \mathcal{D}(\gamma)^2 \log(4/\delta)}{\gamma_A^2 (l+u)}} \tag{4.4.5}$$

since

$$\begin{aligned}
&\left( f_{z,\gamma}(x_i) - f_{z,\gamma}(x_j)^2 \right)^2 - \left( f_\gamma(x_i) - f_\gamma(x_j) \right)^2 \\
&= \left[ \frac{1}{l+u} \sum_{t=1}^{l+u} \left( g_\gamma(x_t) K(x_i, x_t) - g_\gamma(x_t) K(x_j, x_t) \right) \right. \\
&\quad \left. - \int_X \left( g_\gamma(x') K(x_i, x') - g_\gamma(x') K(x_j, x') \right) \mathrm{d}\rho_X(x') \right] \\
&\quad \times \left[ \frac{1}{l+u} \sum_{t=1}^{l+u} \left( g_\gamma(x_t) K(x_i, x_t) - g_\gamma(x_t) K(x_j, x_t) \right) \right. \\
&\quad \left. + \int_X \left( g_\gamma(x') K(x_i, x') - g_\gamma(x') K(x_j, x') \right) \mathrm{d}\rho_X(x') \right] \\
&\le 2\kappa^2 \frac{\mathcal{D}(\gamma)}{\gamma_A} \sqrt{\frac{2\log(2/\delta)}{l+u}} \cdot 4\kappa^2 \frac{\mathcal{D}(\gamma)}{\gamma_A} \\
&\le 8\kappa^2 \sqrt{2\log(2/\delta)} \frac{\mathcal{D}(\gamma)^2}{\gamma_A^2} (l+u)^{-1/2}
\end{aligned} \tag{4.4.6}$$

Then the bound of the following is derived with $\gamma_I = \gamma_A^{2-s}$

$$\frac{\gamma_I}{(l+u)^2}\hat{f}_{z,\gamma}^{\mathrm{T}} L \hat{f}_{z,\gamma} - \frac{\gamma_I}{(l+u)^2} f_\gamma^{\mathrm{T}} L f_\gamma \le 8\kappa^2 \sqrt{2\log(4/\delta)}\mathcal{D}(\gamma)^2 \gamma_A^{-s}(l+u)^{-1/2} \tag{4.4.7}$$

Finally, we have

$$\mathcal{H}_1 \le \kappa \frac{\mathcal{D}(\gamma)}{\sqrt{l+u}}\left\{\frac{2\log(4/\delta)}{\sqrt{l+u}} + \sqrt{2\log(4/\delta)}\right\} + 8\kappa^4\sqrt{2\log(4/\delta)}\frac{\mathcal{D}(\gamma)^2}{\gamma_A^s\sqrt{l+u}} \tag{4.4.8}$$

The $\mathcal{H}_2$ has been proved in [8].□

### 4.5. Estimation of the Sample Error

Since $\hat{f}_{z,\gamma}$ is a function valued random variable which depends on the sample error in the data independent space $\mathcal{H}_\infty$ which contains all possible hypothesis spaces $\mathcal{H}_{K,z}$. Our estimations for $\mathcal{H}_1, \mathcal{H}_2$ are based on the following concentration inequality see [8].

**Lemma 4.2.** *Let $\mathcal{F}$ be a class of measurable function on $Z$. Assume that there are constants $B, c > 0$ and $\beta \in [0,1]$ such that $\|f\|_\infty \le B$ and $Ef^2 \le c(Ef)^\beta$ for every $f \in \mathcal{F}$. If for some $a > 0$ and $\mu \in (0,2)$,*

$$\log \mathcal{N}_2(\mathcal{F}, \zeta) \le a\zeta^{-\mu}, \forall \zeta > 0 \tag{4.5.1}$$

then there exists a constant $c_\mu$ depending only on $\mu$ such that for any $0 < \delta < 1$, with confidence $1-\delta$, there holds

$$Ef - \frac{1}{l}\sum_{i=1}^{l} f(z_i) \le \frac{1}{2}w^{1-\beta}(Ef)^\beta + c_\mu w + 2\left(\frac{c\log(1/\delta)}{l}\right)^{1/(2-\beta)} + \frac{18B\log(1/\delta)}{l}, \forall f \in \mathcal{F} \tag{4.5.2}$$

where $w = \max\left\{c^{\frac{2-\mu}{4-2\beta}}\left(\frac{a}{l}\right)^{\frac{2}{4-2\beta+\mu\beta}}, B^{\frac{2-\mu}{2+\mu}}\left(\frac{a}{l}\right)^{\frac{2}{2+\mu}}\right\}$. The same bound also holds true for $\frac{1}{l}\sum_{i=1}^{l} f(z_i) - Ef$.

The following proposition which has been proved in [9] will be utilized to bound $\mathcal{S}_1$.

**Proposition 4.5.** *suppose that $\rho$ has a τ-quantile of p-average type $q$ for some $p \in (0,\infty]$, $q \in (1,\infty)$. Let $r \le 1$ and $0 < \gamma_A \le 1$. Assume $B_1$ satisfies the capacity assumption* (3.12) *with some $0 < \mu < 2$. Then, for any $0 < \delta < 1$, with confidence $1-\delta$, there holds, for all $f \in B_r$,*

$$\left\{\varepsilon^\tau(\pi(f)) - \varepsilon^\tau(f_\rho^\tau)\right\} - \left\{\varepsilon_z^\tau(\pi(f)) - \varepsilon_z^\tau(f_\rho^\tau)\right\} \le \frac{1}{2}C_1^{1-\theta} r^{\frac{2\mu(1-\theta)}{2+\mu}} l^{\frac{2(1-\theta)}{4-2\theta+\mu\theta}}\left\{\varepsilon^\tau(\pi(f)) - \varepsilon^\tau(f_\rho^\tau)\right\}^\theta + \left(36 + 2C_\theta^{\frac{1}{2-\theta}}\right)\log(1/\delta)\, l^{\frac{-1}{2-\theta}} + C_2 r^{\frac{2\mu}{2+\mu}} l^{\frac{2}{4-2\theta+\mu\theta}} \tag{4.5.3}$$

Here $C_1$ and $C_2$ are the constants depending on $\mu, \theta, c_{\mu,K}$ and $C_\theta$.

The following proposition which has been proved in [9] will be utilized to bound $\mathcal{S}_2$.

**Proposition 4.6.** *Under the assumptions of proposition* 4.5. *Then, for any* $0<\delta<0$, *with confidence* $1-\delta$, *there holds,*

$$\mathcal{S}_2 \le C_3\left(1+\frac{1}{l}\log\frac{5}{\delta}\right)\log\left(\frac{10}{\delta}\right)\times l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}} \times\left(\gamma_A^{s-1}(l+u)^{-\frac{1}{2}}+\gamma_A^{s-1+\theta}(l+u)^{\frac{1-\theta}{2}}\right) \tag{4.5.4}$$

here $C_3$ is a constant independent of $l, \gamma_A, \delta$.

*Proof.* We consider the following function set with $r \ge 1$ to bound $\mathcal{S}_2$

$$\mathcal{G}_r = \left\{\rho_\tau\left(f(x)-y\right)-\rho_\tau\left(f_\rho^{tau}(x)-y\right): f\in B_r\right\} \tag{4.5.5}$$

since $\left|f_\rho^\tau\right| \le 1$ and $\|f\|_\infty \le \kappa r$, for any $g\in\mathcal{G}_r$, we have

$$\left|g(z)\right| \le \left|f(x)-f_\rho^\tau(x)\right| \le \|f\|_\infty + 1 \le \kappa r+1 \tag{4.5.6}$$

By Lemma 3.1, the variance-expectation condition of $g(z)$ is satisfied with $\theta$ given by (3.4) and $c=C_\theta, \beta=\theta$. Then we get

$$\log\mathcal{N}_2\left(\mathcal{G}_r,\eta\right) \le c_{\mu,K} r^\mu \eta^{-\mu}. \tag{4.5.7}$$

Applying lemma 4.2 to $\mathcal{G}_r$, then for any $\delta\in(0,1)$, with confidence $1-\delta$, there holds that, for any $f\in B_r$,

$$\begin{aligned}&\left\{\varepsilon_z^\tau(f)-\varepsilon_z^\tau\left(f_\rho^\tau\right)\right\}-\left\{\varepsilon^\tau(f)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\}\\ &\le \frac{1}{2}\upsilon^{1-\theta}\left\{\varepsilon^\tau(f)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\}^\theta + 2\left(\frac{C_\theta\log(1/\delta)}{l}\right)^{\frac{1}{2-\theta}}\\ &+\frac{18(\kappa r+1)\log(1/\delta)}{l}+c_\mu\upsilon.\end{aligned} \tag{4.5.8}$$

where $\upsilon=\tilde{C} r l^{-\frac{2}{4-2\theta+\mu\theta}}$ and $\tilde{C}=C_\theta^{\frac{2-\mu}{4-2\theta+\mu\theta}} c_{\mu,K}^{\frac{2}{4-2\theta+\mu\theta}}+(\kappa+1)^{\frac{2-\mu}{2+\mu}} c_{\mu,K}^{\frac{2}{2+\mu}}$. From the processing of estimating $\mathcal{H}_1$, for any $\delta\in(0,1)$, with confidence $1-2\delta/5$, we have

$$\frac{1}{l+u}\sum_{i=1}^{l+u}\left|g_\gamma(x_i)\right|-\left\|g_\gamma\right\|_{L^1_{\rho X}} \le \kappa\frac{\mathcal{D}(\gamma)}{\gamma_A\sqrt{l+u}}\left(\frac{2\log(5/\delta)}{\sqrt{l+u}}+\sqrt{2\log(5/\delta)}\right) \tag{4.5.9}$$

which implies there exists a subset $V_1$ of $X^{l+u}$ with measure at most $2\delta/5$ such that

$$\begin{aligned}\frac{1}{l+u}\sum_{i=1}^{l+u} g_\gamma(x_i) &\le \max\left\{\kappa\frac{\mathcal{D}(\gamma)}{\gamma_A\sqrt{l+u}}\left(\frac{2\log(5/\delta)}{\sqrt{l+u}}+\sqrt{2\log(5/\delta)}\right),1\right\}\\ &\triangleq r_\gamma, \forall z\in X^{l+u}\setminus V_1\end{aligned} \tag{4.5.10}$$

The above inequality guarantees $\hat{f}_{z,\gamma}\in B_r$ with for every $x\in X^{l+u}\setminus V_1$. By Lemma 4.2 and (4.5.8), there existing $V_{r_\gamma}$ with measure at most $\delta/5$ such that for every $x \in X^{l+u} \backslash \left(V_1 \cup V_{r_\gamma}\right)$, we have $\hat{f}_{z,\gamma} \in B_{r_\gamma}$ and

$$\begin{aligned}
&\left\{\varepsilon_z^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon_z^\tau\left(f_\rho^\tau\right)\right\}-\left\{\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\} \\
&\le \frac{1}{2}\tilde{C}^{1-\theta} r_\gamma^{1-\theta} l^{\frac{2(1-\theta)}{4-2\theta+\mu\theta}}\left\{\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\rho^{tau}\right)\right\}^\theta \\
&\quad +18(\kappa+1) r_\gamma l^{-1}\log\frac{5}{\delta}+c_\mu\tilde{C} r_\gamma l^{\frac{2}{4-2\theta+\mu\theta}}+2\left(\frac{C_\theta \log(5/\delta)}{l}\right)^{\frac{1}{2-\theta}} \\
&\le \frac{1}{2}\tilde{C}^{1-\theta} r_\gamma^{1-\theta} l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}\left|\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\gamma\right)\right|^\theta \\
&\quad +\frac{1}{2}\tilde{C}^{1-\theta} r_\gamma^{1-\theta} l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}\left\{\varepsilon^\tau\left(f_\gamma\right)-\varepsilon^\tau\left(f_\rho^\tau\right)\right\}^\theta \\
&\quad +18(\kappa+1) r_\gamma l^{-1}\log\frac{5}{\delta}+c_\mu\tilde{C} r_\gamma l^{-\frac{2}{4-2\theta+\mu\theta}}+2\left(\frac{C_\theta \log(5/\delta)}{l}\right)^{\frac{1}{2-\theta}}
\end{aligned} \tag{4.5.11}$$

Proposition 4.2 implies that $\varepsilon^\tau\left(f_\gamma\right)-\varepsilon^\tau\left(f_\rho^\tau\right) \le \mathcal{D}(\gamma) \le C_0\gamma_A^s$ and

$$r_\gamma \le (\kappa+1)\frac{\gamma_A^{s-1}}{\sqrt{l+u}}\left(\frac{2\log(5/\delta)}{\sqrt{l+u}}+\sqrt{2\log(5/\delta)}+1\right) \tag{4.5.12}$$

The Proposition 4.4 tells that there exists a subset $V_2$ of $X^{l+u}$ with measure of at most $2\delta/5$ such that for every $x \in X^{l+u} \backslash V_2$,

$$\varepsilon^\tau\left(\hat{f}_{z,\gamma}\right)-\varepsilon^\tau\left(f_\gamma\right) \le \kappa^2\frac{\mathcal{D}(\gamma)}{\gamma_A\sqrt{l+u}}\left\{\frac{2\log(5/\delta)}{\sqrt{l+u}}+\sqrt{2\log(5/\delta)}\right\} \tag{4.5.13}$$

Let $V = V_1 \cup V_2 \cup V_{r_\gamma}$. Obviously, the measure of $V$ is at most $\delta$ and for every $x \in X^{l+u} \backslash V$, the above inequalities hold. Finally, we combines (4.5.11), (4.5.12), (4.5.13), the result is completed.□

## 5. Total Error Bound

**Proposition 5.1.** *suppose that* $\rho$ *has a* $\tau$*-quantile of* $p$*-average type* $q$ *for some* $p \in (0,+\infty]$ *and* $q \in (1,\infty)$, *and that Approximation condition* (3.7) *and Capacity condition* (3.12) *hold. Let* $0 < \gamma_A \le 1, r \ge 1$ *and* $0 < \delta < 1$. *Then, there exists a subset* $U_r$ *of* $X^{l+u}$ *with measure at most* $\delta$ *such that for any* $x \in W(r)/U_r$, *we have*

$$\begin{aligned}
&\varepsilon^\tau\left(\pi\left(f_{z,\gamma}\right)\right)-\varepsilon^\tau\left(f_\rho^\tau\right)+\gamma_A\Omega\left(f_{z,\gamma}\right)+\frac{\gamma_I}{(l+u)^2} f_{z,\gamma}^{\mathrm{T}} L f_{z,\gamma} \\
&\le \hat{C} r^{\frac{2\mu}{2+\mu}} l^{-\frac{2}{4+\mu\theta-2\theta}}+C_4\left(\frac{2\log(10/\delta)}{\sqrt{l+u}}+\sqrt{2\log(10/\delta)}+1\right)\Psi(l,u,\gamma)
\end{aligned} \tag{5.1}$$

Here $\hat{C}, C_4$ are constants independent of $l, u, \gamma_A, \delta$ and

$$\Psi(l,u,\gamma)=\gamma_A^s+\gamma_A^s(l+u)^{-1/2}+\gamma_A^{s-1}(l+u)^{-1/2}+\gamma_A^{s-1}(l+u)^{-1/2} l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}$$
$$+\gamma_A^{s-1+\theta}(l+u)^{-(1-\theta)/2} l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}$$

*Proof.* Proposition 4.4 ensures the existence $U_1$ of $X^{l+u}$ with measure at most $2\delta/5$ such that

$$\mathcal{H}_1 \le \kappa \frac{\mathcal{D}(\gamma)}{\sqrt{l+u}} \left\{ \frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} \right\} + 8\kappa^4 \sqrt{2\log(10/\delta)} \frac{\mathcal{D}(\gamma)^2}{\gamma_A^s \sqrt{l+u}} \quad (5.2)$$

$$\mathcal{H}_2 \le \kappa^2 \frac{\mathcal{D}(\gamma)}{\gamma_A \sqrt{l+u}} \left( \frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} \right) \quad (5.3)$$

hold for any $x \in X^{l+u}/U_1$.

Proposition 4.5 tell us that there exists a subset $V_r$ of $X^{l+u}$ with measure at most $\delta/10$, such that

$$\begin{aligned} \mathcal{S}_1 \le & \frac{1}{2} C_1^{1-\theta} r^{\frac{2\mu(1-\theta)}{2+\mu}} l^{\frac{2(1-\theta)}{4-2\theta+\mu\theta}} \left\{ \varepsilon^\tau \left( \pi\left( f_{z,\gamma} \right) \right) - \varepsilon^\tau \left( f_\rho^\tau \right) \right\}^\theta \\ & + \left( 36 + 2C_\theta^{\frac{1}{2-\theta}} \right) \log\left( \frac{10}{\delta} \right) l^{-\frac{1}{2-\theta}} + C_2 r^{\frac{2\mu}{2+\mu}} l^{-\frac{2}{4-2\theta+\mu\theta}} \end{aligned} \quad (5.4)$$

Proposition 4.6 ensures the existence of a subset $U_2$ of $X^{l+u}$ with measure at most $\delta/2$ such that

$$\begin{aligned} \mathcal{S}_2 \le & C_3 \left( \frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1 \right) \times l^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}} \\ & \times \left( \gamma_A^{s-1} (l+u)^{-\frac{1}{2}} + \gamma_A^{s-1+\theta} (l+u)^{-\frac{1-\theta}{2}} \right), \ \forall x \in X^{l+u}/U_2 \end{aligned} \quad (5.5)$$

Proposition 4.3 ensures that there exists a subset $U_3$ of $X^{l+u}$ with measure almost $10/\delta$ such that

$$\mathcal{M}(z,\gamma) \le 4\omega\kappa^4 C_0^2 \sqrt{2\log(10/\delta)} \gamma_A^s (l+u)^{-1/2} \quad (5.6)$$

Takeing $U_r = U_1 \cup U_2 \cup U_3 \cup V_r$, the measure of $U_r$ is at most $\delta$, combining (5.2)-(5.6) and (4.2.1), then for every $x \in W(r)/U_r$ we get

$$\begin{aligned} & \varepsilon^\tau \left( \pi\left( f_{z,\gamma} \right) \right) - \varepsilon^\tau \left( f_\rho^\tau \right) + \gamma_A \Omega\left( f_{z,\gamma} \right) + \frac{\gamma_I}{(l+u)^2} f_{z,\gamma}^{\mathrm{T}} L f_{z,\gamma} \\ & + \frac{1}{2} C_4 \left( \frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1 \right) \Psi(l,u,\gamma) \\ & + \frac{1}{2} C_1^{1-\theta} r^{\frac{2\mu(1-\theta)}{2+\mu}} l^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}} \left\{ \varepsilon^\tau \left( \pi\left( f_{z,\gamma} \right) \right) - \varepsilon^\tau \left( f_\rho^\tau \right) \right\}^\theta \\ & + C_2 r^{\frac{2\mu}{2+\mu}} l^{-\frac{2}{4+\mu\theta-2\theta}} \end{aligned} \quad (5.7)$$

Here $C_1, C_2, C_3, C_4, C_5$ is a constant independent of $l, u, \gamma_A, \delta$, and $\Psi(l,u,\gamma)$ are as above.

Next, let $t = \varepsilon^\tau \left( \pi\left( f_{z,\gamma} \right) \right) - \varepsilon^\tau \left( f_\rho^\tau \right) + \gamma_A \Omega\left( f_{z,\gamma} \right) + \frac{\gamma_I}{(l+u)^2} f_{z,\gamma}^{\mathrm{T}} L f_{z,\gamma}$. Hence, the inequality (5.6) can be expressed as

$$t - \frac{1}{2} C_1^{1-\theta} r^{\frac{2\mu(1-\theta)}{2+\mu}} l^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}} t^\theta - \Pi \le 0, \quad (5.8)$$

where $\Pi$ is the rest terms. From Lemma 7.2 in [learning theory: an approximation theory viewpoint], the (5.8) has a unique positive solution $t^*$ which can be bounds as

$$t^* \le \max\left\{C_1 r^{\frac{2\mu}{2+\mu}} l^{\frac{2}{4+\mu\theta_2\theta}}, 2\Pi\right\} \le C_1 r^{\frac{2\mu}{2+\mu}} l^{-\frac{2}{4+\mu\theta-2\theta}} + 2\Pi, \tag{5.9}$$

then the result is derived.□

## 6. Convergence Radius and Main Result

**Proposition 6.1.** *Under the assumptions in proposition* 5.1, *we take* $\omega_0 = \frac{2}{4-2\theta+\mu\theta}$, $\gamma_A = l^{-\beta}$ *with* $\beta > 0$. *Then, for any* $0<\delta<1$, *with confidence* $1-\delta$, *there holds*

$$\begin{aligned}\left\| f_{z,\gamma} \right\| \le &\left( \left(1+\hat{C}\right)^{\frac{2+\mu}{2-\mu}} + \left(N_0+1\right)\hat{C}_4 \right.\\ &\left. \times \left( \frac{2\log\left(10/\delta\right)}{\sqrt{l+u}} + \sqrt{2\log\left(10/\delta\right)+1} \right) \right) \times l^{\beta(1-s)}\end{aligned} \tag{6.1}$$

*Proof.* Applying $\gamma_A = l^{-\beta}$ with $\beta>0$ and letting $\Delta = \frac{2\mu}{2+\mu}$ to proposition 5.1, then for any $r\ge 1$, there exists a subset $V_r$ of $X^{l+u}$ with measure at most $\delta$ such that

$$\left\| f_{z,\gamma} \right\| \le a_m r^{\Delta} + b_m, \forall x \in W\left(r\right)/V_r \tag{6.2}$$

where the constants are given

$$\begin{aligned} a_m &= \hat{C} l^{\beta-\omega_0} \\ b_m &= \hat{C}_4 \left( \frac{2\log\left(10/\delta\right)}{\sqrt{l+u}} + \sqrt{2\log\left(10/\delta\right)+1} \right) \Psi\left(l,u,\gamma\right)/\gamma_A \\ &\triangleq b_\delta \Psi\left(l,u,\gamma\right)/\gamma_A \end{aligned} \tag{6.3}$$

where $\hat{C}, \hat{C}_5$ is a constant independent of $l, u, \gamma_A, \delta$.

It follows that

$$W\left(r\right) \subset W\left(a_m r^{\Delta} + b_m\right) \cup V_r. \tag{6.4}$$

Then, we define a sequence $\left\{r^{(n)}\right\}_{n=1}^{N}$ by $r^{(0)} = \gamma_A^{-1}$ and, for $n \ge 1$

$$r^{(n)} = a_m \left(r^{(n-1)}\right)^{\Delta} + b_m, n \in \mathbb{N}. \tag{6.5}$$

Duo to $\left\| f_{z,\gamma} \right\| \le 1/\gamma_A$ ensures $W\left(r^{(0)}\right) = X^{l+u}$, we have

$$X^{l+u} = W\left(r^{(0)}\right) \subseteq W\left(r^{(1)}\right) \cup V_{r^{(0)}} \subseteq \cdots \subseteq W\left(r^{(N)}\right) \cup \left( \bigcup_{n=0}^{N-1} V_{r^{(n)}} \right). \tag{6.6}$$

with $\rho\left(\bigcup_{n=0}^{N-1} V_{r^{(n)}}\right) \le N\delta$. Hence the measure of is at least. By the iteration formula (6.5), we have

$$\begin{aligned}
r^{(N)} &\le a_m^{1+\Delta+\Delta^2+\cdots+\Delta^{N-1}} \left(r^{(0)}\right)^{\Delta^N} + \sum_{n=1}^{N-1} a_m^{1+\Delta+\Delta^2+\cdots+\Delta^{n-1}} b_m^{\Delta^n} + b_m \\
&= a_m^{\frac{1-\Delta^N}{1-\Delta}} \gamma_A^{-\Delta^N} + \sum_{n=1}^{N-1} a_m^{\frac{1-\Delta^n}{1-\Delta}} b_m^{\Delta^n} + b_m \\
&\le \left(1+\hat{C}\right)^{\frac{1}{1-\Delta}} l^{(\beta-\omega_0)\frac{1}{1-\Delta}+\Delta^N\left(\beta-\frac{1}{1-\Delta}(\beta-\omega_0)\right)} + a_m^{\frac{1}{1-\Delta}} \sum_{n=1}^{N-1}\left(a_m^{\frac{-1}{1-\Delta}} b_m\right)^{\Delta^n} + b_m \qquad (6.7)\\
&\le \left(1+\hat{C}\right)^{\frac{2+\mu}{2-\mu}} l^{(\beta-\omega_0)\frac{2+\mu}{2-\mu}+\Delta^N\left(\frac{2+\mu}{2-\mu}\omega-\frac{2\beta\mu}{2-\mu}\right)} + N a_m^{\frac{1}{1-\Delta}} \max\left\{a_m^{\frac{-1}{1-\Delta}} b_m, 1\right\} + b_m \\
&\le \left(1+\hat{C}\right)^{\frac{2+\mu}{2-\mu}} l^{\beta-\omega_0\frac{2-\mu}{2+\mu}+\Delta^N\left(\frac{2+\mu}{2\mu}\omega_0-\frac{2\beta\mu}{2-\mu}\right)} + (N+1) b_m + N\hat{C}^{\frac{2+\mu}{2-\mu}} l^{(\beta-\omega_0)\frac{2+\mu}{2-\mu}}
\end{aligned}$$

where

$$\begin{aligned}
b_m \le \hat{C}_4 &\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\Big(l^{\beta(1-s)} + l^{\beta(2-s)-1/2} + l^{\beta(1-s)}(l+u)^{-1/2} \\
&+ l^{\beta(2-s)-\omega_0(1-\theta)}(l+u)^{-1/2} + l^{\beta(2-s-\theta)-\omega_0(1-\theta)}(l+u)^{-(1-\theta)/2}\Big)
\end{aligned}$$

Noting that $0<\beta\le\frac{1}{2}<\omega_0$, to ensure that

$$(\beta-\omega_0)\frac{2+\mu}{2-\mu} + \Delta^N\left(\frac{2+\mu}{2-\mu}\omega_0 - \frac{2\beta\mu}{2-\mu}\right) \le \beta(1-s) \qquad (6.8)$$

we only need

$$\Delta^{-N} \ge \frac{\omega_0-\beta\Delta}{\omega_0-\Delta\beta-(1-\Delta)s\beta} \qquad (6.9)$$

Then we get

$$N \le \max\left\{\log_{\frac{2+\mu}{2\mu}} \frac{\omega_0-\beta\Delta}{\omega_0-\Delta\beta-(1-\Delta)s\beta}, 1\right\} \le \log_{\frac{2+\mu}{2\mu}} \frac{\omega_0-\Delta/2}{\omega_0-1/2} + 1 \triangleq N_0 \qquad (6.10)$$

Combine (6.2) and (6.10), we have

$$\begin{aligned}
\left\|f_{z,\gamma}\right\| \le &\Bigg(\left(1+\hat{C}\right)^{\frac{2+\mu}{2-\mu}} + (N_0+1)\hat{C}_4 \\
&\times\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\Bigg) l^{\beta(1-s)},
\end{aligned} \qquad (6.11)$$

with

$$\begin{aligned}
r^{N_0} \le &\Bigg(\left(1+\hat{C}\right)^{\frac{2+\mu}{2-\mu}} + (N_0+1)\hat{C}_4 \\
&\times\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\Bigg) l^{\beta(1-s)}
\end{aligned} \qquad (6.12)$$

where The bound follows by replacing $\delta$ by $\frac{\delta}{N_0}$. □

**Theorem 6.1.** *Assume* (3.7) *and* (3.13) *hold. Taking* $\gamma_A = l^{-\beta}, 0<\beta\le\frac{1}{2}, l=u$

*and* $\gamma_l = \gamma_A^{2-s}$. *Suppose that* $\rho$ *has a* $\tau$*-quantile of* $p$ *average type* $q$ *for some* $p \in (0, +\infty]$ *and* $q \in (1, \infty)$, $p^* = \frac{pq}{p+1} > 0$. *Then for any* $0 < \delta < 1$, *with confidence* $1-\delta$, *we have*

$$
\begin{aligned}
&\left\|\pi\left(f_{z,\gamma}\right) - f_\rho^\tau\right\|_{L_{\rho_X}^{p^*}}^q \\
&\le a\left(\left(1+\hat{C}\right)^{\frac{2\mu}{2+\mu}} + \left(N_0+1\right)\hat{C}_4 \times \left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\right) \\
&\times \left(l^{-\left(\omega_0(1-\theta) - \frac{2\mu}{2+\mu}\beta(1-s)\right)} + l^{-\beta s}\right) + b\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right) \\
&\times \left(l^{\beta s}(l+u)^{-1/2} + l^{-(\omega_0(1-\theta)-\beta(1-s))}(l+u)^{-1/2} \right. \\
&\left. + l^{-(\omega_0(1-\theta)-\beta(1-s-\theta))}(l+u)^{-(1-\theta)/2}\right),
\end{aligned}
\tag{6.13}
$$

*Proof.* Applying Lemma 3.1, proposition 6.1 and proposition 5.1, with confidence $1-\delta$, we have

$$
\begin{aligned}
&\left\|\pi\left(f_{z,\gamma}\right) - f_\rho^\tau\right\|_{L_{\rho_X}^{p^*}}^q \\
&\le C_{q,\rho}\hat{C}\left(\left(1+\hat{C}\right)^{\frac{2\mu}{2+\mu}} + \left(N_0+1\right)\hat{C}_4\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\right)^{\frac{2\mu}{2+\mu}} \\
&\times l^{-\left(\omega_0(1-\theta) - \frac{2\mu}{2+\mu}\beta(1-s)\right)} + C_{q,\rho}C_4\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\Psi_2(l,u,\gamma) \\
&\le a\left(\left(1+\hat{C}\right)^{\frac{2\mu}{2+\mu}} + \left(N_0+1\right)\hat{C}_4\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right)\right) \\
&\times \left(l^{-\left(\omega_0(1-\theta) - \frac{2\mu}{2+\mu}\beta(1-s)\right)} + l^{-\beta s}\right) + b\left(\frac{2\log(10/\delta)}{\sqrt{l+u}} + \sqrt{2\log(10/\delta)} + 1\right) \\
&\times \left(l^{\beta s}(l+u)^{-1/2} + l^{-(\omega_0(1-\theta)-\beta(1-s))}(l+u)^{-1/2} \right. \\
&\left. + l^{-(\omega_0(1-\theta)-\beta(1-s-\theta))}(l+u)^{-(1-\theta)/2}\right)
\end{aligned}
\tag{6.14}
$$

Here $a, b$ is a constant independent of $l, u, \delta$ and

$$
\begin{aligned}
\Psi_2(l,u,\gamma) = {} & l^{-\beta s} + l^{\beta s}(l+u)^{-1/2} + l^{-(\omega_0(1-\theta)-\beta(1-s))}(l+u)^{-1/2} \\
& + l^{-(\omega_0(1-\theta)-\beta(1-s-\theta))}(l+u)^{-(1-\theta)/2}
\end{aligned}
\tag{6.15}
$$

with $\beta \le \frac{1}{2}$ The proof is complete.□

## 7. The Sparsity of the Algorithm

In this subsection, we consider the purpose of investigating sparsity of algorithm (2.7). Here the sparsity means the vanishing of some coefficients in the expansion

$$\hat{f}_{z,\gamma} = \sum_{i=1}^{l+u} \alpha_i K(x, x_i) \tag{7.1}$$

We provide a general result for the vanishing of the coefficient.

**Proposition 7.1.** *Let* $|f_{z,\gamma}| \le 1$, $j \in \{1, \cdots, l+u\}$ *and* $\hat{\alpha} = (\alpha_1, \cdots, \alpha_{l+u})$ *be the coefficient vector of* $\hat{f}_{z,\gamma}$. *If*

$$\max \tau, 1 - \tau\kappa \le \frac{\gamma_A}{8} \tag{7.2}$$

$$\frac{\omega \gamma_I}{2(l+u)^2} \sum_{i,t=1}^{l+u} |\alpha_j| \left( K(x_j, x_i) - K(x_j, x_t) \right)^2 \le \frac{\gamma_A}{8} \tag{7.3}$$

$$\frac{\omega \gamma_I}{(l+u)^2} \sum_{i,t=1}^{l+u} \left| K(x_j, x_i) - K(x_j, x_t) \right| \left| f(x_i) - f(x_t) \right| \le \frac{\gamma_A}{8} \tag{7.4}$$

Then we have $\alpha_j = 0$.

*Proof.* Define the function $F(\alpha)$ in (2.7) to be optimized with $\hat{\alpha} = (\alpha_1, \cdots, \alpha_{j-1}, 0, \cdots, \alpha_{l+u}) \in \mathbb{R}^{l+u}$. Denote $\tilde{\alpha} = (\alpha_1, \cdots, \alpha_{j-1}, 0, \cdots, \alpha_{l+u})$ by substituting the *j*th component of $\hat{\alpha}$ to zero. Comparing $F(\hat{\alpha})$ with $F(\tilde{\alpha})$, since $0 \le K(x_i, x_j) \le \kappa$ and $0 \le \omega_{ij} \le \omega$, we have

$$\begin{aligned} F(\hat{\alpha}) - F(\tilde{\alpha}) \ge{}& (\gamma - \max \tau, 1 - \tau\kappa) |\alpha_j| \\ & - \left( \frac{\omega \gamma_I}{2(l+u)^2} \sum_{i,t=1}^{l+u} |\alpha_j| \left( K(x_j, x_i) - K(x_j, x_t) \right)^2 \right) |\alpha_j| \\ & - \left( \frac{\omega \gamma_I}{(l+u)^2} \sum_{i,t=1}^{l+u} \left| K(x_j, x_i) - K(x_j, x_t) \right| \left| f(x_i) - f(x_t) \right| \right) |\alpha_j| \end{aligned} \tag{7.5}$$

If (7.2)-(7.4) are satisfied, we see from $F(\hat{\alpha}) - F(\tilde{\alpha}) \le 0$ that we must have $\alpha_j = 0$. In order to estimate error $\left\| \hat{f}_{z,\gamma} - f_\rho^\tau \right\|_{L^{p^*}_{\rho_X}}$, we only need to bounds $\varepsilon^\tau(\hat{f}_{z,\gamma}) - \varepsilon^\tau(f_\rho^\tau)$. We thus derive the following inequality, which plays an important role in our mathematical analysis.□

## 8. Conclusion

In this paper, we have a discussion of the lowest convergence rate of quantile regression with manifold regularization optimizing the intrinsic structure using the unlabeled data. The main result is to establish an upper bound for the total error showing less than $O\left( l^{-\left( \frac{1}{2} - \beta \right)} \right)$. Meanwhile, the quantile regression provides a piecewice linearity but a convex technique to overcome difficulties such as a high nonlinearity dependence on the predictor and linear suboptimal models. Finally, the sparsity is analysised in the $l_1$ space.

## Acknowledgements

We would like to thank the referees for their constructive comments and suggestions which have improved the paper. This work was supported by the [the Natural Sciences Foundation of China] under Grant [number 71401124].

## References

[1] Heagerty, P. and Pepe, M. (1999) Semiparametric Estimation of Regression Quantiles with Application to Standardizing Weight for Height and Age in U.S. Children. *Journal of the Royal Statistical Society, Series C*, **48**, 533-551. https://doi.org/10.1111/1467-9876.00170

[2] Koenker, R. and Geling, O. (2001) Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. *Journal of the American Statistical Association*, **96**, 458-468. https://doi.org/10.1198/016214501753168172

[3] Koenker, R. and Hallock, K. (2001) Quantile Regression: An Introduction. *Journal of Economic Perspectives*, **15**, 43-56. https://doi.org/10.1257/jep.15.4.143

[4] Shi, L., Huang, X., Tian, Z. and Suykens, J.A.K. (2013) Quantile Regression with Regularization and Gaussian Kernels. *Advances in Computational Mathematics*, **40**, 517-551.

[5] Felipe, C. and Ding, Z. (2007) Learning Theory: An Approximation Theory Viewpoint. Cambridge Monographs on Applied and Computational Mathematics. www.cambridge.org/9780521865593

[6] Joachims (1999) Transductive Inference for Text Classification Using Support Vector Machines. *Proceedings of the Sixteen International Conference on Machine Learning*, 200-209.

[7] Bousquet, O., Chapelle, O. and Hein, M. (1999) Measure Based Regularization. *Asvances in Neural Information Processing Systems*, **16**.

[8] Li, M., Zhang, M. and Sun, H. (2015) Conditional Qunantile Regression with $l_1$ Regularization and Instensitive Pinball Loss. *International Journal of Wavelets, Multiresolution and Information Processing*, **13**.

[9] Li, M. and Hong, W.S. (2015) Asymptotic Analysis of Quantile Regression Learning Based on Coefficient Dependent Regularization. *International Journal of Wavelets, Multiresolution and Information Processing*, **13**, Article ID: 1550018.